

518 Rec'd PCT/PTO 24 AUG 2001

FORM PTO-1390

U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE

4121-129

U.S. APPLICATION NO. (If known, see 37 CFR 1.5)

**09/914549**

TRANSMITTAL LETTER TO THE UNITED STATES  
DESIGNATED/ELECTED OFFICE (DO/EO/US)  
CONCERNING A FILING UNDER 35 U.S.C. 371

INTERNATIONAL APPLICATION NO.

PCT/DE00/00583

INTERNATIONAL FILING DATE

28 February 2000

PRIORITY DATE CLAIMED

26 February 1999

TITLE OF INVENTION

PROTEIN (TP) THAT IS INVOLVED IN THE DEVELOPMENT OF THE NERVOUS SYSTEM

APPLICANT(S) FOR DO/EO/US

Annemarie Poustka and Johannes Coy

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

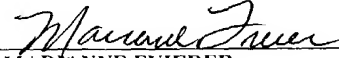
1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☐ This express request to begin national examination procedures (35 U.S.C. 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) and PCT Articles 22 and 39(1).
4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
  - a. ☒ is transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☒ has been transmitted by the International Bureau.
  - c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☒ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. ☐ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
  - a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
  - b. ☐ have been transmitted by the International Bureau.
  - c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
  - d. ☒ have not been made and will not be made.
8. ☐ A translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).\*(Unsigned)
10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

## Items 11. to 16. below concern other document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
13. ☒ A FIRST preliminary amendment.  
☐ A SECOND or SUBSEQUENT preliminary amendment.
14. ☐ A substitute specification.
15. ☒ A small entity statement.
16. ☐ Other items or information: EPO Search Report and International Preliminary Examination Report in German, Computer Readable Disk with Sequence Listing

NOTE: This application is being filed with an unsigned Oath or Declaration under the provisions of 37 CFR § 1.53 in order that applicant may secure a filing date of August 24, 2001. Upon receipt of a "Notice to File Missing Parts - Filing Date Granted," a executed Declaration and Power of Attorney will be forwarded. The undersigned agent affirmatively states that she has been duly authorized and appointed to file this application on behalf of the applicants and applicants' assignee, and that the Declaration and Power of Attorney to be filed hereafter will confirm the undersigned agent's authorization and appointment. Applicants are considered a small entity and assignee Deutsches Krebsforschungszentrum is also considered a small entity within the meaning of 37 CFR § 1.9.

*Handwritten:* Filing Date Granted  
10/1/01

17. <input checked="" type="checkbox"/> The following fees are submitted:				CALCULATIONS	PTO USE ONLY
<b>Basic National Fee (37 CFR 1.492(a)(1)-(5)):</b> Search Report has been prepared by the EPO or JPO .....\$860.00  International preliminary examination fee paid to USPTO (37 CFR 1.482) .....\$0.00 No International preliminary examination fee paid to USPTO (37 CFR 1.482) but international search fee paid to USPTO (37 CFR 1.445(a)(2)) .....\$0.00  Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO .....\$1000.00  International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(2)-(4) .....\$0.00  <b>ENTER APPROPRIATE BASIC FEE AMOUNT =</b>				\$ 860.00	
Surcharge of \$130.00 for furnishing the oath or declaration later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				\$	
Claims	Number Filed	Number Extra	Rate		
Total Claims	38-20 =	18	X \$18.00	\$ 324.00	
Independent Claims	12-3 =	9	X \$80.00	\$ 720.00	
Multiple dependent claim(s) (if applicable)			+ \$000.00	\$	
<b>TOTAL OF ABOVE CALCULATIONS =</b>				1904.00	
Reduction by 1/2 for filing by small entity, if applicable. Verified Small Entity statement must also be filed. (Note 37 CFR 1.9, 1.27, 1.28).				\$ 952.00	
<b>SUBTOTAL =</b>				\$ 952.00	
Processing fee of \$130.00 for furnishing the English translation later than <input type="checkbox"/> 20 <input type="checkbox"/> 30 Months from the earliest claimed priority date (37 CFR 1.492(f)).				\$	
<b>TOTAL NATIONAL FEE =</b>				\$ 952.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). \$40.00 per property				\$	
<b>TOTAL FEE ENCLOSED =</b>				\$ 430.00	
				Amount to be: refunded	\$
				Charged	\$522.00
a. <input checked="" type="checkbox"/> A check in the amount of \$430.00 to cover part of the above fees is enclosed. b. <input checked="" type="checkbox"/> Please charge my Deposit Account No. 08-3284 in the amount of \$522.00 to cover the above fees. A duplicate copy of this sheet is enclosed. c. <input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 08-3284. A duplicate copy of this sheet is enclosed.					
<b>NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not yet been met, a petition to revive (37 CFR 1.127(a) or (b)) must be filed and granted to restore the application to pending status.</b>					
SEND ALL CORRESPONDENCE TO:  <b>Steven J. Hultquist</b> <b>Intellectual Property/Technology Law</b> <b>P. O. Box 14329</b> <b>Research Triangle Park, NC 27709</b>					
<div style="text-align: right;">   <b>MARIANNE FUIERER</b>  <b>Registration No. 39,983</b> </div>					

Rec'd PCT/PTO

04 JAN 2002

4549 0050502

4121-129

PATENT APPLICATION

---

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

---

**In re Application of:** Poustka, et al.  
**Application No.:** New U.S. National Stage Application of  
PCT International Application No.  
PCT/DE00/00583  
**International Filing Date:** 28 February 2000  
**Priority Date Claimed:** 26 February 1999 (German Appl. No. 199 048  
423.8)  
**U.S. National Phase Filing Date:** Date of mailing identified below  
**Title:** **PROTEIN (TP) THAT IS INVOLVED IN  
THE DEVELOPMENT OF THE NERVOUS  
SYSTEM**

---

**EXPRESS MAIL CERTIFICATE**

I hereby certify that I am mailing the attached documents to the  
Commissioner for Patents on the date specified, in an envelope  
addressed to the Commissioner for Patents, Box Patent Application,  
Washington, DC 20231, and Express Mailed under the provisions of  
37 CFR 1.10.

Blake Crouch

Name of Person Mailing This Document

*Blake Crouch*

Signature

August 24, 2001

Date

EL831358276US

Express Mail Label Number

---

**PRELIMINARY AMENDMENT**

---

Commissioner for Patents  
BOX PATENT APPLICATION  
Washington, D.C. 20231

Sir:

### In the Claims

1. A DNA sequence coding for a protein which is involved in the development of the nervous system, in particular the CNS, and is expressed in a tissue-specific and development-specific manner, wherein the DNA sequence comprises the following DNA sequences:

- 2



DNA sequence of (a), (b), (c), (d), (e), (f), (g), (h) or (i); or

- (k) a DNA sequence which differs from the DNA sequence of (a), (b), (c), (d), (e), (f), (g), (h), (i) or (j) due to the degeneration of the genetic code.
2. The DNA sequence according to claim 1, which codes for a protein or peptide comprising the amino acid sequence of figure 1, figure 9, figure 11, figure 12, figure 13, figure 14, figure 15, figure 16, figure 17, figure 18 or figure 19, wherein the protein or peptide has the biological activity defined in claim 1.
  3. An antisense RNA, characterized in that it is complementary to the DNA sequence of claim 1 and can reduce or inhibit the synthesis of the protein encoded by this DNA sequence.
  4. A ribozyme, characterized in that it is complementary to the DNA sequence of claim 1 and can bind specifically to the RNA transcribed by this DNA sequence and can cleave it so as to reduce or inhibit the synthesis of the protein encoded by this DNA sequence.
  5. An expression vector, containing the DNA sequence selected from the group consisting of the protein according to claim 1 the antisense RNA according to claim 3 or the ribozyme according to claim 4.
  6. The expression vector according to claim 5, which comprises additionally the promoter of the human T gene or an ortholog of the T gene.
  7. An expression vector according to claim 6, which codes for a protein selected from the group consisting of T, T2, T3 proteins or for fragments thereof in the form of a reporter fusion protein.
  8. A host cell which is transformed with an expression vector selected from the

9. A protein which is encoded by the DNA sequence according to claim 1 and which is involved in the development of the nervous system and is expressed in tissue-specific and development-specific manner, or fusion proteins, fragments, variants, derivatives or precursors of the protein

Motive 1:

### Motive 2:

### Motive 3:

### Motive 4:

Motive 5:

### Motive 6:

### Motive 7:

Xs at this site.

12. Antibody which is directed against the protein according to claim 9 or fragment

thereof.

13. Antibody according to claim 12, which is obtained by immunizing animals with a peptide having the sequence  
"EKGEDPETRRMRTVKNIADI".
14. A method for preventing or treating diseases of the nervous system by using a member selected from the group consisting of the DNA sequence according to claim 1, the antisense RNA according to claim 3, the ribozyme according to claim 4, the expression vector according to any of claims 5 to 7, the protein according to claim 9 and the antibody or the fragment thereof according to claim 12 or 13 for preventing or treating diseases of the nervous system, in particular of the CNS.
15. The method according to claim 14, wherein the disease of the nervous system is a tumoral disease of the CNS.
16. The method according to claim 14, wherein the treatment of diseases of the nervous system are the promotion of the neuronal regeneration in the case of injuries of the nervous system and degenerative diseases of the nervous system.
17. The method according to claim 14, wherein the treatment of diseases of the nervous system are the regeneration of the neuronal linkages and the regeneration of the innate and acquired malfunctions of the nervous system.
18. The method according to claim 15 for inhibiting the growth and spreading of tumor cells.
19. Diagnostic method for detecting a disturbed expression of the protein according to claim 9 or for detecting a changed form of this protein, in which a sample is contacted with a member selected from the group consisting of the DNA sequence according to claim 1, the DNA sequence according to claim 2, the

antibody or the fragment thereof according to claim 12, and the antibody or the fragment thereof of claim 13 and then it is determined directly or indirectly whether the concentration of the protein and/or its amino acid sequence differs from a protein obtained from a healthy patient.

20. Diagnostic kit for carrying out the method according to claim 19, which contains at least one member selected from the group consisting of the DNA sequence according to claim 1, the DNA sequence according to claim 2, the antibody or the fragment thereof according to claim 12, and the antibody or the fragment thereof according to claim 13.
21. Non-human mammal whose naturally occurring T, T2 or T3 gene comprises a change in the gene structure or the gene sequence.
22. Non-human mammal, wherein a change of the gene structure of the T, T2 or G3 gene is achieved in the mammal by introducing a deletion in place of which a homologous or heterologous sequence is introduced.
23. Non-human mammal, wherein a change of the gene structure of the T, T2 or C3 gene is achieved by inserting a homologous or heterologous sequence in the corresponding gene naturally occurring in the mammal.
24. Non-human mammal according to claim 22, wherein the heterologous sequence is the selection marker sequence.
25. Non-human mammal according to claim 24, wherein the selection marker sequence conveys resistance to neomycin.
26. A method of producing a non-human mammal selected from the group consisting of the non-human mammal according to claim 21, claim 22, claim 23, claim 24 and claim 25, characterized by the steps of:
  - (a) producing a DNA fragment, in particular a vector, containing a changed

- T, T2 or G3 gene, the T, T2 or T3 gene having been modified by inserting a heterologous sequence, in particular a selectable marker;
- (b) preparing embryonal stem cells from a non-human mammal (preferably a mouse);
  - (c) transforming the embryonal stem cells from step (b) with the DNA fragment from step (a), the T gene in the embryonal stem cells being changed by homologous recombination with the DNA fragment from (a)
  - (d) culturing the cells from step (c),
  - (e) selecting the cultured cells from step (d) for the presence of the heterologous sequence, in particular the selectable marker,
  - (f) producing chimeric non-human mammals from the cells of step (e) by injecting these cells into mammalian blastocysts (preferably mouse blastocysts), transferring the blastocysts to pseudo-pregnant female mammals (preferably mouse) and analyzing the resulting offspring for a change of the T, T2 or T3 gene.
27. Transgenic cell or tissue which is capable of expressing a T protein or part of the T protein or an ortholog thereof.
28. A method for the analysis of the function of the T gene family by using a member selected from the group consisting of the non-human mammal according to claim 21, claim 22, claim 23, claim 24, claim 25 the transgenic cell of claim 27 or the transgenic tissue according to claim 27.
29. A method for identifying inhibitors and enhancers of the T gene family by using the non-human mammal according to claim 21, claim 22, claim 23, claim 24, claim 25, the transgenic cell according to claim 27 or the transgenic tissue according to claim 27.
30. Vertebrate gene and functional equivalent, derivative or a bioprecursor thereof, which code for a protein having a statistically significant amino acid sequence homology to the T gene, T2 gene or T3 gene according to any of the following

figures selected from the group consisting of: figure 1, figure 9, figure 11, figure 12, figure 13, figure 14, figure 15, figure 16, figure 17, figure 18 or figure 19.

31. T gene and its vertebrate orthologs and vertebrate paralogs which code for a nuclear pore protein.
32. Vertebrate protein which has an amino acid sequence according to figure 1 or an amino acid sequence which differs from the amino acid sequence in figure 1 by one or more amino acids.
33. Vertebrate T, T2 or T3 gene and the protein encoded therein in all of its naturally occurring allelic and mutated forms.
34. Medicament containing a protein according to claim 9 or a functional equivalent, a fragment or a bioprecursor thereof in combination with a pharmaceutically acceptable carrier.
35. The method of identifying substances which has an enhancing or inhibiting influence on the effect of T protein, T2 protein or T3 protein, by means of
  - determining the bi-directional transport through the nuclear pores,
  - determining the binding to filaments of the cell (e.g. actin filaments and microtubuli) or
  - determining the increased or reduced transcription of cellular or reporter genes.
36. Method of identifying substances which have an enhancing or inhibiting influence on the effect of proteins which are functionally linked to the T protein in direct or indirect way, or represent parallel signal or functional pathways, by means of
  - determining the bi-directional transport through the nuclear pores,

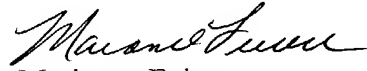
- determining the phosphorylation and the dephosphorylation of proteins,
  - determining the binding of the T protein to filaments of the cell (e.g. actin filaments and microtubuli), or
  - determining the increased or reduced transcription of cellular or reporter genes.
37. The method according to claim 35, wherein the modified transcription with reporter molecules, preferably the occurrence of certain mRNAs or the EGEP protein, is detected.
38. The method of identifying further proteins which play a role in the development and function of the nervous system and/or are a nuclear pore protein, wherein the method comprises the steps of:
- (a) producing an antibody against a protein according to claim 9,
  - (b) contacting a cell extract with the antibody and identifying the antibody/protein complex,
  - (c) analyzing the complex to identify a protein which has bound to the protein of the complex and is no antibody, and
  - (d) optionally repeating steps (a) to (c) to identify further proteins of this function.

**REMARKS**

A marked-up version of amended paragraph in the specification and amended claims 1-38 are included herewith in Appendix A.

It is requested that the examination and prosecution of this application proceed on the basis of the English translation of the PCT International application included herewith and these amended claims 1-38.

Respectfully submitted,



Marianne Fuierer

Registration No. 39,983

Attorney for Applicants

INTELLECTUAL PROPERTY/  
TECHNOLOGY LAW  
P. O. Box 14329  
Research Triangle Park, NC 27709  
Phone: (919) 419-9350  
Fax: (919) 419-9354  
Attorney File: 4121-129



## APPENDIX A

1. A DNA sequence coding for a protein which is involved in the development of the nervous system, in particular the CNS, and is expressed in a tissue-specific and development-specific manner, wherein the DNA sequence comprises the following DNA sequences:

- (a) the DNA sequence of figure 1, figure 2, figure 3, figure 4, figure 5, figure 6, figure 7 or figure 8;
- (b) the DNA sequence of figure 9 or figure 10;
- (c) the DNA sequence of figure 11;
- (d) the DNA sequence of figure 12 or figure 13;
- (e) the DNA sequence of figure 14 or figure 15;
- (f) the DNA sequence of figure 16;
- (g) the DNA sequence of figure 17 or 18;
- (h) the DNA sequence of figure 19;
- (i) a DNA sequence hybridizing with (a), (b), (c), (d), (e), (f), (g) or (h)
- (j) fragments, variants, functional equivalents, derivatives or precursors of the DNA sequence of (a), (b), (c), (d), (e), (f), (g), (h) or (i); or
- (k) a DNA sequence which differs from the DNA sequence of (a), (b), (c), (d), (e), (f), (g), (h), (i) or (j) due to the degeneration of the genetic code.

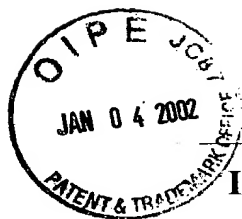
3. An antisense RNA, characterized in that it is complementary to the DNA sequence of claim 1 [or 2] and can reduce or inhibit the synthesis of the protein encoded by this DNA sequence.
4. A ribozyme [Ribozyme], characterized in that it is complementary to the DNA sequence of claim 1 [or 2] and can bind specifically to the RNA transcribed by this DNA sequence and can cleave it so as to reduce or inhibit the synthesis of the protein encoded by this DNA sequence.
5. An expression [Expression] vector, containing the DNA sequence selected from the group consisting of the protein according to claim 1 [or 2 or coding for] the antisense RNA according to claim 3 or the ribozyme according to claim 4.
7. An expression [Expression] vector according to claim [5 or] 6, which codes for a protein selected from the group consisting of [for the] T, T2, [or] T3 proteins or for fragments thereof in the form of a reporter fusion protein.
8. A host [Host] cell which is transformed with [the] an expression vector selected from the group consisting of the expression vector of claim 5, claim 6 and claim 7. [according to any of claims 5 to 7.]
9. A protein [Protein] which is encoded by the DNA sequence according to claim 1 [or 2] and which is involved in the development of the nervous system and is expressed in tissue-specific and development-specific manner, or fusion proteins, fragments, variants, derivatives or precursors of the protein
11. A method [Method] of producing the protein according to claim 9, which comprises culturing the host cell according to claim 8 under suitable conditions and obtaining the protein from the cell or the culture medium.

14. A method for preventing or treating diseases of the nervous system by using a member selected from the group consisting of [Use of ]the DNA sequence according to claim 1 [or 2], the antisense RNA according to claim 3, the ribozyme according to claim 4, the expression vector according to any of claims 5 to 7, the protein according to claim 9 and [or] the antibody or the fragment thereof according to claim 12 or 13 for preventing or treating diseases of the nervous system, in particular of the CNS.
15. The method [Use] according to claim 14, wherein the disease of the nervous system is a tumoral disease of the CNS.
16. The method [Use] according to claim 14, wherein the treatment of diseases of the nervous system are the promotion of the neuronal regeneration in the case of injuries of the nervous system and degenerative diseases of the nervous system.
17. The method [Use] according to claim 14, wherein the treatment of diseases of the nervous system are the regeneration of the neuronal linkages and the regeneration of the innate and acquired malfunctions of the nervous system.
18. The method [Use] according to claim 15 for inhibiting the growth and spreading of tumor cells.
19. Diagnostic method for detecting a disturbed expression of the protein according to claim 9 or for detecting a changed form of this protein, in which a sample is contacted with a member selected from the group consisting of the DNA sequence according to claim 1. the DNA sequence according to claim 2, [or 2 or] the antibody or the fragment thereof according to claim 12, and the antibody or the fragment thereof of claim [or] 13 and then it is determined directly or indirectly whether the concentration of the protein and/or its amino acid sequence differs from a protein obtained from a healthy patient.

20. Diagnostic kit for carrying out the method according to claim 19, which contains at least one member selected from the group consisting of the DNA sequence according to claim 1, the DNA sequence according to claim [or] 2, [and/or] the antibody or the fragment thereof according to claim 12, and the antibody or the fragment thereof according to claim [or] 13.
24. Non-human mammal according to claim 22 [or 23], wherein the heterologous sequence is the selection marker sequence.
26. A method of producing a non-human mammal selected from the group consisting of the non-human mammal according to claim 21, claim 22, claim 23, claim 24 and claim 25, [to any of claims 21 to 25], characterized by the steps of:
- (a) producing a DNA fragment, in particular a vector, containing a changed T, T2 or G3 gene, the T, T2 or T3 gene having been modified by inserting a heterologous sequence, in particular a selectable marker;
  - (b) preparing embryonal stem cells from a non-human mammal (preferably a mouse);
  - (c) transforming the embryonal stem cells from step (b) with the DNA fragment from step (a), the T gene in the embryonal stem cells being changed by homologous recombination with the DNA fragment from (a)
  - (d) culturing the cells from step (c),
  - (e) selecting the cultured cells from step (d) for the presence of the heterologous sequence, in particular the selectable marker,
  - (f) producing chimeric non-human mammals from the cells of step (e) by injecting these cells into mammalian blastocysts (preferably mouse blastocysts), transferring the blastocysts to pseudo-pregnant female mammals (preferably mouse) and analyzing the resulting offspring for a change of the T, T2 or T3 gene.
28. A method for the analysis of the function of the T gene family by using a member selected from the group consisting of the [Use of the] non-human mammal

according to [any of claims] claim 21, claim 22, claim 23, claim 24, claim [to] 25  
or] the transgenic cell of claim 27 or the transgenic tissue according to claim 27.  
[for the analysis of the function of the T gene family.]

29. A method for identifying inhibitors and enhancers of the T gene family by using  
[Use of] the non-human mammal according to claim 21, claim 22, claim 23, claim  
24, claim 25, [to any of claims 21 to 25 or] the transgenic cell according to claim  
27 or the transgenic tissue according to claim 27. [for identifying inhibitors and  
enhancers of the T gene family.]
30. Vertebrate gene and functional equivalent, derivative or a bioprecursor thereof,  
which code for a protein having a statistically significant amino acid sequence  
homology to the T gene, T2 gene or T3 gene according to any of the following  
figures selected from the group consisting of: figure 1, figure 9, figure 11, figure  
12, figure 13, figure 14, figure 15, figure 16, figure 17, figure 18 or figure 19.
37. The method according to claim 35 [or 36], wherein the modified transcription  
with reporter molecules, preferably the occurrence of certain mRNAs or the  
EGEP protein, is detected.



09914549 .060500  
JC05 Rec'd PCT/PTO , 0 4 JAN 2002

4121-129  
PATENT APPLICATION

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

**In re Application of:**

**Poustka, et al.**

**Application No.:**

**09/914,549**

**International Application No.:**

**PCT/DE00/00583**

**Priority Date Claimed:**

**28 February 2000 and 26 February 1999  
(German Appl. No. 199 048 423.8)**

**Title:**

**PROTEIN (TP) THAT IS INVOLVED IN THE  
DEVELOPMENT OF THE NERVOUS  
SYSTEM**



**23448**

PATENT TRADEMARK OFFICE

**FIRST CLASS MAIL CERTIFICATE**

I hereby certify that I am mailing the attached documents to the Commissioner for Patents on the date specified, in an envelope addressed to the Commissioner for Patents, Washington, DC 20231, and First Class Mailed under the provisions of 37 CFR 1.8.

*Lee Ann Brown*

Lee Ann Brown

November 14, 2001

Date of Mailing

**SECOND SUPPLEMENTAL PRELIMINARY AMENDMENT IN U.S. PATENT  
APPLICATION NO. 09/914,549**

Commissioner for Patents  
Washington, D.C. 20231

Sir:

Prior to examination of the above-identified national phase patent application, please amend the application, as follows:

**In the Specification**

Please insert on page 1 between the title of the application and the first paragraph the following new paragraph:

**CROSS-REFERENCE TO RELATED APPLICATIONS**

This application is filed under the provisions of 35 U. S.C. §371 and claims the priority of International Patent Application No. PCT/DE00/00583 filed February 28, 2000, which in turn claims priority of German Patent Application No. 199 048 423.8 filed on February 26, 1999.

**REMARKS**

This claim to priority is being filed before the above-identified application meets all the requirements under 35 U.S.C. §371(b).

Respectfully submitted,



Marianne Fuierer  
Registration No. 39,983  
Attorney for Applicants

INTELLECTUAL PROPERTY/  
TECHNOLOGY LAW  
P. O. Box 14329  
Research Triangle Park, NC 27709  
Phone: (919) 419-9350  
Fax: (919) 419-9354  
Attorney File: 4121-129

K 3008

124/ppts

Protein (TP) That is Involved in the Development of the  
Nervous System

The present invention relates to a protein (T protein) and to proteins related thereto which are involved in the development of the nervous system and are expressed in a tissue-specific and development-specific manner, to the below described variants of these proteins and to DNA sequences coding for these proteins. The present invention further relates to antibodies directed against these proteins or to fragments thereof as well as to antisense RNAs or ribozymes directed against the expression of these proteins. Finally, the present invention concerns medicaments and diagnostic methods in which the above-mentioned compounds are used.

Mutations in genes playing a part in the development and maintenance of the nervous system are of utmost scientific and economic significance, since diseases of the nervous system, in particular CNS, occur frequently, are often characterized by a severe, partly fatal disease process and can be treated only to a limited extent thus far. The increase in the life expectancy is accompanied by a drastic increase in neurological and psychic diseases. The latter greatly limit the quality of life of the affected persons and cause considerable costs for both the affected person and the public.

Isolating and analyzing genes specific to the nervous system offer a good possibility of studying diseases, such as schizophrenia, Alzheimer's disease, autism, manic depression



and mental retardation, and eventually of also being able to treat them.

The present invention is thus based on the technical problem of providing products by means of which disturbances in the development and function of the nervous system can be diagnosed and optionally be treated.

The solution to this technical problem is achieved by providing the embodiments characterized in the claims.

The subject matter of the present invention is thus a DNA sequence coding for a protein which is involved in the development and function of the nervous system, in particular the CNS, and is expressed in tissue-specific and development-specific manner, the DNA sequence comprising the following DNA sequences:

- (a) the DNA sequence of figure 1, figure 2, figure 3, figure 4, figure 5, figure 6, figure 7 or figure 8;
- (b) the DNA sequence of figure 9 or figure 10;
- (c) the DNA sequence of figure 11;
- (d) the DNA sequence of figure 12 or figure 13;
- (e) the DNA sequence of figure 14 or figure 15;
- (f) the DNA sequence of figure 16;
- (g) the DNA sequence of figure 17 or 18;
- (h) the DNA sequence of figure 19;
- (i) a DNA sequence hybridizing with (a), (b), (c), (d), (e), (f), (g) or (h);
- (j) variants, derivatives, precursors or fragments of the DNA sequence of (a), (b), (c), (d), (e), (f), (g), (h) or (i); or

- (k) a DNA sequence differing from the DNA sequence of (a), (b), (c), (d), (e), (f), (g), (h), (i) or (j) due to the degeneration of the genetic code.

The present invention is based on the isolation of a human DNA sequence (referred to as gene "T" or T gene; see figures 1 to 8, which codes for protein TP), it turning out that the protein encoded by this DNA sequence is required in the nervous system. In this connection, the expression of the gene encoding this protein is increased in the nervous system. A sequence analysis showed that it is a new gene. Moreover, further genes could be isolated which have homologies to this gene (murine gene "T", figures 9 and 10; human gene "T2", figure 16; human gene "T3", figures 17 and 18; murine gene T2, figures 12 and 13; murine gene T3, figure 19). The T gene, T2 gene and T3 gene are members of the T (gene) family, as shown below, and originate preferably from vertebrates, such as man, mouse or rat. Defects in these genes limit the functions of the nervous system, in particular the CNS. These genes also perform an important function in the control of cell growth, and changes in these genes or their expression result in defects regarding the control of cell growth, e.g. also in tumor formation, in particular of the neuroblastoma. Small children up to the age of 8 are affected almost exclusively by this cancerous disease. The first symptoms already occur within the first 12 months of life in 25 to 30 percent of the cases. In the case of the neuroblastoma very young cells of the autonomous nervous system degenerate. Since these nerves extend along the rear side of the abdominal region and the chest, neuroblastomas usually occur in the regions of the stomach, pelvis, chest and neck. More than half the diseases start from the suprarenal marrow which is also formed by nerve cells. Symptoms which may refer in small

children to a neuroblastoma are nodes, swellings, bone pain, limping, tiredness, fever, paleness, sweating, obstinate or persistent cough, hematomas around the eye. A neuroblastoma can be diagnosed by a physician by means of blood tests, urine analyses and ultrasonic examinations and by the removal of biopsies from the tumor and an examination of bone marrow. As soon as the accurate location of the tumor is diagnosed, it is removed by means of an operation. However, the early formation of metastases creates a problem. By isolating and analyzing the T gene it is now possible to develop novel measures of diagnosing and treating the neuroblastoma. Due to this, it is possible to diagnose the cancerous disease early and establish forms of therapy promising better chances of recovery.

Mutations in genes of the T gene family also lead to a disturbed development and differentiation of the nervous system, in particular the brain. In many cases, this results in mental diseases, e.g. mental retardations or Alzheimer's disease. The T gene also plays an important role in the interconnection of individual regions of the brain, e.g. forebrain and midbrain. Mutations in this gene lead in some cases to schizophrenic diseases and syndromes of autism. By means of the human and murine genes it is possible to draw important fundamental conclusions as to the development of the nervous system and in particular the brain. Good approaches offer themselves as regards the research of pathologic changes of the nervous system and in particular the brain.

Patients can be examined more simply for possible mutations by means of the genomic sequences. The genomic sequences of the T gene are of advantage in particular when little (tumor) material is available for the analysis. By this it

is possible, for example, to examine even minute tumors for mutations in this gene. This also provides the possibility of checking a therapy (in particular radiation therapy and/or chemotherapy) for its being successful, since it is possible to detect tumor cells circulating in the blood by genomic primers which are specific to the genomic DNA using a PCR reaction.

The term "hybridizing" used in the present invention relates to conventional hybridization conditions, preferably to hybridization conditions which use 5xSSPE, 1 % SDS, 1xDenhardt's solution as the solution and where hybridization temperatures are between 35°C and 70°C, preferably 65°C. Following hybridization, washing is preferably carried out using first 2xSSC, 1 % SDS and then 0.2xSSC at temperatures between 35°C and 70°C, preferably of 65°C (regarding a definition for SSPE, SSC and Denhardt's solution see Sambrook et al., Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). Stringent hybridization conditions are particularly preferred, as described in Sambrook et al., *supra*, for example.

The terms "variants" or "fragment" used in the present invention comprise DNA sequences which differ from the sequences indicated in the figures by deletion(s), insertion(s), substitution(s) and/or other modifications known in the art or comprise a fragment of the original nucleic acid molecule, the protein or peptide encoded by these DNA sequences still having the above-mentioned properties. Therefore, functional equivalents, derivatives, precursors (bioprecursors) are counted among them. Derivatives are understood to mean e.g. mutation derivatives

(produced by deletions or insertions, for example), fusions, allele variants, muteins and splicing variants. Two select examples of such splicing variants are shown in figures 14 and 15. Methods of producing the above changes in the nucleic acid sequence are known to a person skilled in the art and are described in standard works of molecular biology, e.g. in Sambrook et al., *supra*. The person skilled in the art is also capable of determining whether a protein encoded by a nucleic acid sequence modified in such a way still has the above-mentioned properties.

In a preferred embodiment, the present invention relates to a DNA sequence which encodes a protein comprising the amino acid sequence of figure 1, figure 9, figure 11, figure 12, figure 13, figure 14, figure 15, figure 16, figure 17, figure 18 or figure 19, the protein having the above-defined biological activity.

By lowering or inhibiting the expression of the above described DNA sequences it is possible to reduce or eliminate the synthesis of the proteins encoded by them, e.g. the T protein, which is desirable for certain states of a disease, for example. Therefore, another preferred embodiment of the present invention relates to antisense RNA, which is characterized in that it is complementary to the above DNA sequences and can reduce or inhibit the synthesis of the protein encoded by these DNA sequences and to a ribozyme, which is characterized in that it can bind specifically to part of the above DNA sequences and to the RNA transcribed by these DNA sequences and can cleave them so as to reduce or inhibit the synthesis of the protein encoded by these DNA sequences. These antisense RNAs and ribozymes are preferably complementary to a coding region of the mRNA. Based on the disclosed DNA sequences, the person

skilled in the art can produce and use suitable antisense RNAs. Suitable methods are described in EP-B1 0 223 399 or EP-A1 0 458, for example. Ribozymes are RNA enzymes and consist of a single RNA strand. They can cleave intermolecularly other RNAs, e.g. the mRNAs transcribed by the DNA sequences according to the invention. These ribozymes must, in principle, have two domains: (1) a catalytic domain and (2) a domain which is complementary to the target RNA and can bind thereto, which is a precondition for a cleavage of the target RNA. Based on the methods described in the literature, it is meanwhile possible to construct specific ribozymes which excise a desired RNA at a certain pre-select site (see e.g. Tanner et al., in: Antisense Research and Applications, CRC Press, Inc. (1993), 415-426).

The DNA sequences according to the invention or the DNAs encoding the above described antisense RNAs or ribozymes may also be inserted in a vector or expression vector. Thus, the present invention also comprises vectors or expression vectors containing these DNA sequences. The term "vector" relates to a plasmid (e.g. pUC18, pBR322, pBlueScript), to a virus or another suitable vehicle. In a preferred embodiment, the DNA molecule according to the invention is functionally linked in the vector to regulatory elements allowing the expression thereof in prokaryotic or eukaryotic host cells. Along with the regulatory elements, e.g. a promoter, such vectors contain typically a replication origin and specific genes which allow the phenotypic selection of a transformed host cell. The lac, trp promoter or the T7 promoter are counted among the regulatory elements for the expression in prokaryotes, e.g. *E. coli*, those for the expression in eukaryotes comprise the AOX1 or GAL1 promoter in yeast, and those for the expression in animal

cells include the CMV, SV40, RVS40 promoter, CMV or SV40 enhancer. Further examples of suitable promoters are the metallothionein I promoter and the polyhedrin promoter. In a preferred embodiment the vector contains the promoter of the human T gene or an ortholog of the T gene. Suitable expression vectors for *E. coli* are e.g. pGEMEX, pUC derivatives, pGEX-2T, pET3b and pQE-8, the latter being preferred. Suitable vectors for the expression in yeast comprise pY100 and Ycpad1, and suitable vectors for the expression in mammalian cells include pMSXND, pKCR, pEFBOS, cDM8 and pCEV4. Vectors derived from baculovirus for expression in insect cells, e.g. pAcSGHisNT-A, are also counted among the expression vectors according to the invention.

General methods known in the art can be used for constructing expression vectors which contain the DNA sequences according to the invention and suitable control sequences. These methods comprise e.g. *in vitro* recombination techniques, synthetic methods, and *in vivo* recombination techniques, as described in Sambrook *et al.*, *supra*, for example. The DNA sequences according to the invention can also be inserted in combination with a DNA coding for another protein or peptide, so that the DNA sequences according to the invention can be expressed in the form of a fusion protein, for example. These other DNAs are preferably reporter sequences which code for a reporter molecule comprising a detectable protein, e.g. a stain or coloring matter, an antibiotic resistance,  $\beta$ -galactosidase or a substances detectable by spectrophotometric, spectrofluorometric, luminescent or radioactive assays.

The present invention also relates to host cells containing the above described vectors. These host cells comprise

bacteria (e.g. the *E. coli* strains HB101, DH1, x1776, JM101, JM109, BL21 and SG13009), fungi, e.g. yeasts, preferably *S. cerevisiae*, plant cells, insect cells, preferably sf9 cells, and animal cells, preferably cells from vertebrates or mammals. Preferred mammalian cells are CHO, VERO, BHK, HeLa, COS, MDCK, 293 or WI38 cells. Methods of transforming these host cells for the phenotypic selection of transformants and for the expression of the DNA molecules according to the invention using the above-described vectors are known in the art.

The genes belonging to the sequences according to the invention can be amplified by suitable primer sequences. The primer sequences indicated in figure 20 are particularly suited for amplification of genes T2 and T3.

The present invention also relates to the proteins encoded by the DNA sequences according to the invention and to methods of producing the protein encoded by the DNA sequences according to the invention. The person skilled in the art is familiar with conditions of culturing transformed or transfected host cells. The method according to the invention comprises the culturing of the above described host cells under conditions which allow the expression of the protein (or fusion protein) (preferably stable expression) and the collection of the protein from the culture or from the host cells. Suitable purification methods (e.g. preparative chromatography, affinity chromatography, e.g. immunoaffinity chromatography, HPLC, etc.) are generally known.

The proteins according to the invention preferably comprise the amino acid sequences shown in figure 1, figure 9, figure 11, figure 12, figure 13, figure 14, figure 15, figure 16,



figure 17, figure 18 or figure 19 or represent fusions, fragments, derivatives or precursors (bioprecursors) thereof, the above mentioned properties being maintained within the meaning of functional equivalents. As to the definitions of these terms, reference is made to the respective explanations above. Derivatives are understood to mean in particular the changed proteins or peptides which differ from the sequences shown in the figures by conservative amino acid substitutions or contain non-conserved amino acid substitutions that do not change the function of the T proteins to a substantial degree.

The following amino acid motives have been identified by Inventors. They are suited to identify formerly unknown proteins which belong to the T/T2/T3 family according to the invention and a protein superfamily from pore membrane proteins and filament-binding proteins.

Motive 1:

(A,T) (I,P,V) (L,T) (G,A,Q) (L,V) XXX (L,V)

Motive 2:

IYTDWAN

Motive 3:

XXXXXXXXXXGXXXXXXXXXXXXXXXXXXXXXXXXXXQ

Motive 4:

SXXXXDX (12,20) KX (17,22) XXXXXXXXXL

Motive 5:

IYTDWANXXLX (K,R)

Motive 6:

KX (18,21) XXXXXXXXXLX (15,24) S

Motive 7:

NX (3,11) SXXXAXXXXXXXL

Explanation: X stands for every amino acid

(A,T) means amino acid A or T at this site

X(2,4) denotes two to four Xs at this site

Another preferred embodiment of the present invention relates to antibodies against the above described proteins according to the invention or to a fragment thereof. These antibodies may be monoclonal, polyclonal or synthetic antibodies or fragments thereof. In this connection, the term "fragment" means all parts of the monoclonal antibody (e.g. Fab, Fv or "single chain Fv" fragments) which have an epitope specificity the same as that of the complete antibody. The person skilled in the art is familiar with the production of such fragments.

The antibodies according to the invention are preferably monoclonal antibodies. The antibodies according to the invention can be produced according to standard methods, the protein encoded by the DNA sequences according to the invention or a synthetic fragment thereof serving as an immunogene. Methods of obtaining monoclonal antibodies are known to the person skilled in the art and comprise e.g. as a first step the production of polyclonal antibodies using the proteins according to the invention or fragments thereof (synthetic peptides, for example) as an immunogene for immunizing suitable animals, e.g. rabbits or chickens, and the collection of the polyclonal antibodies from the serum or egg yolk.

For example, cell hybrids from cells producing antibodies and tumor cells from bone marrow are then produced and cloned. Thereafter, a clone is selected which produces an antibody specific to the antigen used. This antibody is then produced. Examples of cells producing antibodies are spleen cells, lymph node cells, B lymphocytes, etc. Examples of animals which can be immunized for this purpose are mice,

rats, horses, goats and rabbits. The myeloma cells can be obtained from mice, rats, humans or other sources. The cell fusion can be carried out by the generally known method developed by Köhler and Milstein, for example. The hybridomas obtained by cell fusion are screened using the antigen according to the enzyme-antibody method or according to a similar method. Clones are obtained with the boundary dilution method, for example. The resulting clones are implanted intraperitoneally into BALB/c mice, for example, the mouse ascites is removed after 10 to 14 days, and the monoclonal antibody is purified by known methods (e.g. ammonium sulfate fractionation, PEG fractionation, ion exchange chromatography, gel chromatography or affinity chromatography).

In a particularly preferred embodiment, said monoclonal antibody is an antibody originating from an animal (e.g. mouse), a humanized antibody or a chimeric antibody or a fragment thereof. Chimeric antibodies similar to human antibodies or humanized antibodies have a reduced potential antigenicity, however, their affinity is not lowered over the target. The production of chimeric and humanized antibodies or of antibodies similar to human antibodies has been described in detail (see e.g. Queen *et al.*, Proc. Natl. Acad. Sci., U.S.A. 86 (1989), 10029, and Verhoeven *et al.*, Science, 239 (1988), 1534). Humanized immunoglobulins have variable framework regions which originate substantially from a human immunoglobulin (designated acceptor immunoglobulin) and the complementarity of the determining regions which originate substantially from a non-human immunoglobulin (e.g. from a mouse) (designated donor immunoglobulin). The constant region(s) originate(s), if available, also substantially from a human immunoglobulin. When administered to human patients, humanized (and the

human) antibodies have a number of advantages over antibodies from mice or other species: (a) the human immune system should not regard the framework or the constant region of the humanized antibody as foreign and therefore the antibody response to such an injected antibody should be less than to that to a completely foreign mouse antibody of a partially foreign chimeric antibody; (b) since the effector region of the humanized antibody is human, it might interact better with other parts of the human immune system, and (c) injected humanized antibodies have a half life which is substantially equivalent to that of human antibodies occurring in nature, which permits the administration of doses smaller and less frequent as compared to antibodies of other species.

The antibodies according to the invention can be used for the immunoprecipitation of the above discussed proteins, for the isolation of related proteins from cDNA expression libraries or for the below indicated purposes (diagnosis/therapy), for example.

The present invention also relates to a hybridoma which produces the above described monoclonal antibody.

In a preferred embodiment, the present invention relates to antibodies against the peptides of genes T2 and T3 listed separately (*cf.* figure 20).

It has been found that the below peptide can be used specifically for generating antibodies against the T protein. The amino acid sequence of the suitable peptide reads as follows:

EKGEDPETRRMRTVKNIAD

The present invention makes possible to study disturbances in the development and function of the nervous system on a genetic level. These disturbances comprise *inter alia* neurological and psychiatric diseases (*inter alia* Alzheimer's disease, Parkinson's disease, schizophrenia, manic-depressive diseases, autism, mental retardations), injuries of the nervous system, innate damage of the nervous system or degenerative diseases of the nervous system. The invention also enables the treatment of cancer, *inter alia* of tumors of the nervous system, such as neuroblastoma, astrocytoma, glioblastoma, medulloblastoma. This diagnosis cannot only be made postnatally but already prenatally. It can be detected by means of the DNA sequence according to the invention or probes or primers derived therefrom whether mammals, in particular humans, contain a gene which codes for and/or expresses the protein according to the invention or whether this gene results in a mutated form of the protein which is no longer biologically active. For this purpose, the person skilled in the art can carry out common methods, such as reverse transcription, PCR, LCR, hybridization and sequencing. The antibodies according to the invention are also suited e.g. for diagnosis, i.e. for detecting in a sample the presence and/or concentration of the protein according to the invention, a shortened or extended form of the protein, etc. The antibodies can be bound e.g. in immunoassays in liquid phase or to a solid carrier. In this case, the antibodies can be labeled in various ways. Suitable markers and labeling methods are known in the art. Examples of immunoassays are ELISA and RIA.

Thus, the present invention also relates to a diagnostic method for detecting a disturbed expression of the protein

according to the invention or for detecting a changed form of this protein, in which a sample is contacted with the DNA sequences according to the invention or the antibody according to the invention or the fragment thereof and then it is determined directly or indirectly whether the concentration of the protein and/or its amino acid sequence differs from a protein obtained from a healthy patient.

The present invention also allows to carry out therapeutic measures in connection with the above discussed disturbances, i.e. the above described inventive DNA sequences, antisense RNAs, ribozymes and antibodies can also be used for producing a medicament, e.g. for controlling the expression of the protein according to the invention, or for exchanging a mutated form of the gene by a functional form and thus also for the production of a medicament for preventing or treating diseases of the nervous system, in particular tumoral diseases of the CNS. For example, the protein according to the invention can be introduced into mammals, in particular humans, by common measures. For this purpose, it may be favorable to link the protein to a protein which is not considered foreign by the respective body, e.g. transferrin or bovine serum albumin (BSA). An inventive DNA sequence, antisense RNA or ribozyme can also be introduced into mammals, in particular humans, and expressed. By means of an antibody according to the invention it is possible to control and regulate the expression of the protein (TP) according to the invention or the related proteins.

Thus, the present invention also relates to a medicament which contains the above described DNA sequences, antisense RNA, the ribozyme, the expression vector, the protein according to the invention or the antibody or the fragment

thereof. This medicament contains, optionally in addition, a pharmaceutically compatible carrier. Suitable carriers and the formulation of such medicaments are known to the person skilled in the art. Suitable carriers are e.g. phosphate-buffered common salt solutions, water, emulsions, e.g. oil-in-water emulsions, wetting agents, sterile solutions, etc. The medicaments can be administered orally or parenterally. The topical, intra-arterial, intra-muscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal or intranasal administration are counted among the methods for the parenteral administration. The suitable dose is determined by the attending physician and depends on various factors, e.g. on the age, sex and weight of the patient, the stage of the disease, the kind of administration, etc.

The above described nucleic acids are preferably inserted in a vector suitable for gene therapy and introduced into the cells under the control of a tissue-specific vector, for example. In a preferred embodiment, the vector containing the above described nucleic acids is a virus, e.g. an adenovirus, vaccinia virus or adenovirus. Retroviruses are particularly preferred. Examples of suitable retroviruses are MoMuLV, HaMuSV, MuMTV, RSV or GaLV. For the purposes of gene therapy, the nucleic acids according to the invention can also be transported to the target cells in the form of colloidal dispersions. They comprise liposomes or lipoplexes, for example (Mannino *et al.*, *Biotechniques* 6 (1988), 682).

Finally the present invention relates to a diagnostic kit for carrying out the above described diagnostic method, which contains a DNA sequence according to the invention or the above described antibody according to the invention or a

fragment thereof. Depending on the kind of the kit, the DNA sequence or the antibody or the fragment thereof can be immobilized.

Sequences of the T genes can be applied to nylon membranes or glass carriers and hybridized with complex cDNA samples from tumors and pertinent normal tissues or diseased and pertinent healthy tissue. This enables the (fully automated) detection of the expression of these genes. The sequences used for this purpose can be e.g. the entire cDNA sequence or short sequence segments, e.g. 10-15 bp oligomers (see *inter alia* figure 20). Having determined the expression of the T genes, the therapy, *inter alia* the cancer therapy, can be selected deliberately according to the respective individual situation of the patient or can be adapted thereto. Genes whose changed expression influence already now the treatment of the patient are the N-myc gene in the case of neuroblastoma, for example. By detecting the expression of the T genes it is thus possible to adapt the treatment very quickly and efficiently to the respective requirements and in this way it contributes essentially to the improved therapy.

The isolation and characterization of the human gene according to the invention and in particular of the mouse homologues thereof also allow to establish an animal model, which is very valuable for the further study of diseases of the nervous system and of cancerous diseases on a molecular level. The subject matter of the present invention thus also relates to a non-human mammal whose T gene or T2 or T3 gene is changed, e.g. by inserting a heterologous sequence, in particular a selection marker sequence.



The expression "non-human mammal" comprises any mammal whose T gene or T2 or T3 gene can be changed. Examples of such mammals are mouse, rat, rabbit, horse, cattle, sheep, goat, monkey or ape, pig, dog and cat, with mouse being preferred.

The expression "T gene or T2 or T3 gene which is changed" signifies that a change of the gene structure or the gene sequence is carried out by standard methods in the corresponding gene occurring naturally in the non-human mammal. This can be achieved *inter alia* by introducing a deletion of about 1-2 kb, at the place of which a heterologous sequence, e.g. a construct for mediating antibiotic resistance (e.g. a "neo cassette") is introduced. Heterologous sequences allowing to carry out time-specific and tissue-specific deletions *in vivo* can also be inserted in the T gene. Furthermore, heterologous sequences allowing to track the expression of the T gene *in vivo* can be introduced into the T gene. This can be done *inter alia* by inserting a sequence coding for the GFP (green fluorescent protein) protein inside an exon or as an independent exon. These methods are generally described by Schwartzberg *et al.*, Proc. Natl. Acad. Sci., U.S.A., Vol. 87, pages 3210-3214, 1990, to which reference is made herein.

In particular, the modification can be described and carried out as follows. Figure 9 represents part of the cDNA sequence of the T gene of a mouse. Illustration 10 shows an intron sequence of the T gene of a mouse, which is flanked by two exons. These murine sequences can then be used for the deliberate change of the murine T gene. For example, the splicing sequences of the intron can be deleted or changed such that the T gene is no longer spliced correctly. By incorporating a splicing acceptor sequence of another exon of the murine T gene into the intron sequence it is possible

to insert in this intron a sequence which is recognized as exon and is spliced to the T gene exon upstream thereof. This inserted sequence may be an exon, for example, which encodes the EGFP protein (EnhancedGreenFluorescentProtein). As a result, the original murine T gene becomes a fusion protein comprising the EGFP protein. Thus, a mouse can preferably be generated, which allows to track the expression of the T gene *in vivo*. The inserted sequence can be designed at its end (e.g. PolyA signal, splicing signals, etc.) such that no further exons of the T gene are spliced to the inserted exon or the spliced exon can no longer be translated. As a result, a deletion of the murine T protein forms on the C-terminal end or a premature discontinuance of the reading frame, and an (at least partial) inactivation of the protein function of the murine T gene can be achieved. It is also possible to insert, as new exon sequences, sequences which yield an mRNA sequence where this new mRNA sequence is localized at the 3' end. By suitable sequences it is then possible to achieve a change in the stability of the mRNA or a changed localization in the cell. The accompanying phenotypes of the thus modified mice can then result in important conclusions drawn on the function of the T gene. These mice can then also be used for detecting new active substances compensating the functional loss of the T gene.

In another preferred embodiment, the sequence of figure 13 is used for the production of a knock-out mouse. Figure 13 describes a murine sequence of gene T2. The elimination of the murine T2 genes can in this connection be achieved in different ways. For example, the splicing sequence (GT, underlined in figure 13) can be deleted or changed such that the T2 gene is no longer spliced correctly. In addition, by incorporating a splicing acceptor sequence of another exon

of the murine T2 gene into the following intron sequence it is possible to insert in this intron a sequence which is detected as exon and spliced to the T2 gene exons upstream thereof. This inserted exon may be e.g. an exon which codes for the EGFP protein. Due to this, the original murine T2 gene becomes a fusion protein which carries the EGFP protein at the C terminus. In this way, a mouse can be generated which allows to track the expression of the T2 gene *in vivo*. The inserted sequence can be designed at its end (e.g. PolyA signal, etc.) such that no further exons are spliced to the inserted exon by the T2 gene. A deletion of the murine T2 protein forms at the C-terminal end and an (at least partial) inactivation of the protein function of the murine T2 gene can be achieved. Such sequences can also be inserted as new exon sequences which yield an mRNA sequence in which at the 3' end this new mRNA sequence is localized. By means of suitable sequences it is then possible to achieve a change in the stability of the mRNA or a changed localization in the cell. The accompanying phenotypes of the thus changed mice can then lead to important conclusions as to the function of the T2 gene. These mice can also be used for detecting new active substances which compensate the functional loss of the T gene.

Furthermore, a mammal can be generated comprising a change in the T3 gene. The sequence in figure 19 represents part of the murine cDNA sequence of the T3 gene. Deliberate changes in the T3 gene of a mouse can be achieved by deletions or insertions. The inserted sequence can be an exon, for example, which codes for the EGFP protein. As a result, the original murine T3 gene becomes a fusion protein which carries the EGFP protein at the C terminus. Thus, a mouse can be generated which allows to track the expression of the T3 gene *in vivo*. The inserted sequence can be designed at

its end (e.g. PolyA signal, etc.) such that no further exons are spliced from the T3 gene to the inserted exon. A deletion of the murine T3 protein thus forms on the C-terminal end and an (at least partial) inactivation of the protein function of the murine T3 gene can be achieved. It is also possible to insert, as new exon sequences, sequences which yield an mRNA sequence where this new mRNA sequence is localized at the 3' end. By suitable sequences it is then possible to achieve a change in the stability of the mRNA or a changed localization in the cell. The accompanying phenotypes of the mice changed in this way can then lead to important conclusions as to the function of the T3 gene. These mice can then also be used for discovering new active substances which compensate the functional loss of the T3 gene.

Another subject matter of the present invention are cells which are obtained from the above non-human mammal. These cells can be present in any form, e.g. in a primary or long-term culture.

A non-human mammal according to the invention can be provided by common methods. A method is favorable which comprises the steps of:

- (a) producing a DNA fragment, in particular a vector, containing a changed T, T2 or G3 gene, the gene having been modified by inserting a heterologous sequence, in particular a selectable marker;
- (b) preparing embryonal stem cells from a non-human mammal (preferably a mouse);

- (c) transforming the embryonal stem cells from step (b) with the DNA fragment from step (a), the T gene in the embryonal stem cells being changed by homologous recombination with the DNA fragment from (a);
- (d) culturing the cells from step (c);
- (e) selecting the cultured cells from step (d) for the presence of the heterologous sequence, in particular the selectable marker;
- (f) producing chimeric non-human mammals from the cells from step (e) by injecting these cells into mammalian blastocysts (preferably mouse blastocysts), transferring the blastocysts to pseudo-pregnant female mammals (preferably mouse) and analyzing the resulting offspring for a change of the T gene.

In step (c), the mechanism of homologous recombination (*cf.* R.M. Torres, R. Kühn, Laboratory Protocols for Conditional Gene Targeting, Oxford University Press, 1997) is used to transfect embryonal stem cells. The homologous recombination between the DNA sequences present in a chromosome and new added cloned DNA sequences enable the insertion of a cloned gene in the genome of a living cell in place of the original gene. Using embryonal germ cells, animals which are homozygous for the desired gene or the desired gene portion or the desired mutation can be obtained via chimeras by this method.

The expression "embryonal stem cells" comprises any embryonal stem cells of a non-human mammal, suited to mutate the T gene. The embryonal stem cells originate preferably from a mouse, in particular the cells E14/1 or 129/SV.

The expression "vector" comprises any vector which by recombination with the DNA of embryonal stem cells enables a change of the T, T2 or G3 gene. The vector preferably has a marker by means of which selection can be made for existing stem cells in which the desired recombination has been made. Such a marker is the loxP/tk neo cassette, for example, which can be removed by means of the Cre/loxP system from the genome again.

The person skilled in the art also knows conditions and materials serving for carrying out steps (a) - (f).

By means of the present invention a non-human mammal is provided whose T, T2 or T3 gene is changed. This change may be an elimination of the gene expression-regulating function. Using such a mammal or cells therefrom it is possible to study selectively the gene expression-controlling function of the TP protein. It is also possible by this to find substances, medicaments and therapy approaches by which selective influence on the controlled function is possible. The present invention therefore provides a basis for influencing the most differing diseases. Such diseases are e.g. limitations of the CNS functions which cover mental retardations or the induction of cancer resulting from deficiencies in the control of cell proliferation.

Inventors found out in the sequence analysis that the T2 gene in the coding region of the cDNA sequence contains CGG trinucleotides which are known to be sensitive to methylation. The T2 gene thus has in the coding region (N-terminal region of the protein which has no homology to the T protein or T3 protein) a methylation-sensitive and

unstable sequence which results in the failure of the gene accompanied by a mental retardation and uncontrolled cell growth, such as cancer.

All the three genes of the T family play a major role in the case of tumors. The T gene is affected in many tumors by genomic rearrangements. For example, in neuroblastomas genomic changes in the DNA of tumors can be found as compared to the DNA of the accompanying healthy tissue. The expression of the T gene, e.g. in tumors of the brain, is also changed. A strongly changed expression can be found *inter alia* in the advanced stages of glioblastomas. Tumor-specific changes of the expression of the T gene and the occurrence of the T protein can also be detected in meningiomas.

In many tumoral diseases, the T2 gene also undergoes genomic rearrangements, and a changed expression can be detected in tumors. For example, in melanomas and lung tumors genomic rearrangements of the T2 gene can be detected. Expression differences are also detectable in gliomas, glioblastomas, astrocytomas and PNETs (Primitive Neuro-Ectodermal Tumors), for example.

In many tumors, the T3 gene also undergoes genomic rearrangements and expression changes. Rearrangements can be detected in colon carcinomas, for example. Expression differences are detectable *inter alia* in gliomas, glioblastomas, astrocytomas and PNETs (Primitive Neuro-Ectodermal Tumors).

By isolating and accurately analyzing the T gene, Inventors now have found that the T protein has a certain relationship to proteins which perform completely different functions in

the cell. The sequence analysis of these proteins showed that the genes coding for these proteins are likely due to a common precursor gene or to similar precursor genes. Proteins such as the POM121 protein (Hallberg et al., J. Cell Biol. 122, pages 513-522, 1993) belong to this superfamily. It is one of two known nuclear pore membrane proteins in vertebrates. The CLIP-170 protein which binds vesicles and other organelles within the cell to microtubuli (Pierre et al., Cell 70, pages 887-900, 1992) also belongs to this family. The unexpected discovery that genes which perform such different tasks inside the cell belong to a common protein superfamily is extremely surprising and even inconsistent at first sight. However, when the functions of the individual genes are analyzed, two main functions of these proteins can be derived. The CLIP-170 protein binds to microtubuli, the newly isolated T proteins and the POM121 protein are localized in the nuclear core complex. Due to the properties of these proteins, Inventors propose that this protein superfamily be referred to as POMIC protein superfamily. POMIC shall, in this connection, stand for pores and/or microtubuli-binding protein. Based on the isolation and analysis of the T gene, two paralogs of the T gene could be isolated, namely the T2 and T3 genes which are described in more detail above. As regards evolution and function, the family of the T proteins is between the CLIP (cytoplasmic linker protein-170) and the POM121 protein. This intermediate position is also supported by the sequence analysis and the putative protein structure. The nuclear pore membrane protein POM121 has no marked coiled-coil structure whereas the CLIP-170 protein shows a very distinct coiled-coil structure between the N-terminus and C-terminus (cf. figure 29). Coiled-coil structures exist in the family of T proteins, however, they are clearly less marked than in CLIP-170. A similar intermediate position is adopted by the



family of T proteins with respect to the occurrence of hydrophobic domains. The POM121 protein has a hydrophobic domain at the N-terminus which is introduced into the nuclear membrane, and the protein is positioned in the nuclear pore. The CLIP-170 protein has no distinct hydrophobic domain. The T protein and the T3 protein, however, have a hydrophobic domain with three hydrophobic partial regions (cf. figure 30). The exchange of the N-terminus in the T2 protein as compared to the evolutionary basic form resulted in a loss of this distinct hydrophobic domain. Yet all three T proteins have in common the very similar structure of the C-terminus. The T3 protein is most similar to the T protein within the T protein family. However, the T3 protein also has undergone a change in the course of evolution. The N-terminus was changed as compared to the T protein by insertion of about 400 amino acids. This insertion resulted in another coiled-coil structure as compared to the otherwise very similar T protein. The T protein and the T3 protein perform functions in the nuclear membrane-localized form, which are similar to those of POM121. However, it is interesting that in the course of evolution there was a loss of part of the C-terminus in the POM121 protein. As compared to the POM121 protein, the T proteins have a longer C-terminus. Due to this longer C-terminus many interactions with other proteins are possible. In this connection, it is also worth mentioning that a leucine-zipper structure was discovered in the T protein, which facilitates interactions with other proteins. The family of T protein plays an important role in the mediation of interactions between cell organelles and filaments, *inter alia* microtubuli. Microtubuli play an important role e.g. in nerve cells; in the case of axons, for example, the plus ends of the microtubuli face away from the cell body whereas the microtubuli of dendrites have both orientations. This

cell polarity is of major importance for the functioning of a cell or living being. Microtubuli also provide an efficient organelle transport, and they are of essential significance for the general organization of membrane structures in a cell. The T proteins perform an important mediator function between membrane structures and microtubuli. The T gene and the T3 gene perform their function in particular as a membrane protein in the nuclear pore whereas the T2 protein acts particularly as a cytoplasmic protein.

Due to the findings of Inventors the T gene and the T3 gene are part of the nuclear pore complex. Nuclear pore complexes (NPCs) are extremely complicated structures which mediate the bi-directional transport of macromolecules between the nucleus and the cytoplasm. The nuclear pore complex is embedded in the nuclear envelope and encases a central channel with a structure only defined insufficiently thus far. Peripheral structures, short cytoplasmic filaments and a basket-like structure are attached on both sides of the central nuclear pore complex. This basket-like structure interacts with molecules which pass through the nuclear pore complex. The mechanism of synthesizing nuclear pore complexes is hardly understood thus far. In addition, it has been found when observing cells passing through mitosis that the nuclear envelope is dissolved deliberately and their components, including the nuclear pore proteins, are distributed over the mitotic cytoplasm. At the end of mitosis, all these components are used again to form the nuclear envelope of the daughter cells. Due to the detailed analysis of the gene T, Inventors found that the N-terminal half of the T protein is weakly homologous to the pore membrane protein POM121. The homology covers the entire region of the POM121 protein and has an identity of about 18

% on a protein level so that the DNAs underlying these proteins should not hybridize with one another, even under hardly stringent conditions. As regards the formation and structure of the nuclear pore, the T protein according to the invention plays a very fundamental role. In a detailed analysis of the protein, a lipophilic domain could be detected at the N-terminus of the T protein. However, this sequence has no homology to the lipophilic sequence of the POM121 protein. There is also a short segment of amino acids which might serve as a signal sequence before the lipophilic domain in the T protein. In order to find out whether this putative signal sequence and the lipophilic domain are involved *in vivo* in the incorporation into the nuclear membrane, various constructs of the T gene were produced. Various parts of the N-terminus of the T protein were fused with the EnhancedGreenFluorescentProtein (EGFP). The EGFP was here fused to the C-terminus of the T protein. The fusion protein which comprised the unchanged N-terminus of the T protein (putative signal sequence with lipophilic membrane domain) was actually incorporated into the nuclear membrane. However, the fusion construct from which the putative signal sequence and the lipophilic domains lack, was not incorporated into the nuclear membrane and accumulated in the cytoplasm. This showed that the N-terminus of the T protein is necessary and suffices to result in a localization within the nuclear membrane. In order to show that the T protein is actually localized in the nuclear membrane, antibodies were generated against a peptide sequence of the T protein. Immunohistochemical studies of tissues of man, mouse and rat were carried out with these antibodies. It showed that the antibody detects a protein which is localized in the nuclear membrane. Since it is difficult to differentiate by means of a light microscope whether the protein is localized in the nuclear membrane or

the nucleus itself, further analyses were made using the high-resolution method of electron microscopy. By this it was possible to clearly show that the T protein is localized in the nuclear membrane. As a detection reaction a second antibody was used here to which the enzyme horseradish peroxidase was coupled and which resulted in a color reaction (DAB). The stain or coloring formed can be seen in the electron-microscopic pictures only on the cytoplasmic side of the nuclear membrane. This indicates that the antibody recognizes an epitope of the T protein which is accessible from the cytoplasmic side for the antibody. The analysis of the immunohistochemical sections also showed that the antibody recognizes very specific neurons (cf. figure 24). The results of the analysis of the expression on a protein level by means of the antibody are highly consistent with the results of the analysis of the RNA expression. The mouse ortholog of the T gene was used in the RNA *in situ* analyses. Using the human T gene cDNA clones, murine cDNA clones of the mouse ortholog were initially isolated and sequenced for this purpose. The sequence analysis confirmed that the isolated cDNA clones was the mouse ortholog. Such a murine cDNA clone of the T gene was then used for the RNA *in situ* hybridization (cf. figures 25, 26, 27, 28). An expression analysis of the T gene of the mouse was then possible by means of this technique. The accurate analysis of the spatial-temporal expression profile showed that the T gene plays a decisive role in the generation, formation and maintenance of the nervous system in vertebrates. No expression can be detected during the early mouse embryogenesis on day 9.5 *post conceptionem* (pc = *post conceptionem*). On day 10.5 pc, it is possible to detect an expression in the ventral mesencephalon and in the telencephalon. In this stage there is also a strong expression in the connecting region of the mesencephalon and

telencephalon (forebrain-midbrain). An expression of the T gene in the telencephalon, in the ventral mesencephalon and in the myelencephalon can be detected on day 11.5 pc. An expression in neurons of the mantle zone of the developing brain and in the nuclei of the peripheral nerves is visible on day 12.5 pc. Furthermore, there is an expression in the myelencephalon, spinal cord and spinal ganglia. A minor expression is detectable in the mesencephalon and telencephalon. No expression is detectable e.g. in proliferating neurons in the subventricular layer or in the migrating neurons of the 'intermediate' zone. On day 14.5 pc, an expression in mesenchymal tissues, e.g. around the vertebra or in the region of developing bones, is also visible. A strong expression in all parts of the brain and the peripheral nervous system (e.g. spinal ganglia and nerve fibers of the tail) can be detected on day 16.5 pc. An expression in differentiating neurons of the mantle zone of the telecephalons can also be detected. Furthermore, an expression in neurons of the spinal cord and the spinal ganglia can be detected. When the brain develops after the birth, an expression in the olfactory bulb, in the cerebral cortex and in the developing hippocampus can be detected above all. A minor expression is found however in the coliculus and the developing cerebellum. A similar expression pattern exists in the fully developed brain.

Northern blots (*cf.* figure 23) were carried out to find out where the T gene or T2 or T3 gene are expressed. The T gene is expressed predominantly in the brain, hardly or not at all in the heart, lungs, placenta, liver, skeletal muscle, kidney or pancreas (irrespective of adult or fetal tissue). However, the T2 gene is virtually not expressed in the brain but strongly expressed in the heart (adult and fetal), adult liver, adult skeletal muscle and adult kidney. The T3 gene

is expressed in all tested tissues (adult and fetal heart, brain, liver, kidney: placenta, adult skeletal muscle, adult pancreas), except in fetal lungs.

Because of the discovery of the T gene and the detailed analysis of this gene with the information obtained therefrom a basis has been created for the development of fully novel medicaments and medicament compound classes. The bi-directional transport of molecules through the nuclear membrane is of decisive significance for the function of each eukaryotic cell. The information which is stored in the form of DNA (chromosomes) in the nucleus is transcribed into mRNA. However, the information is only translated into protein in the cytoplasm. If the transcribed information (mRNA) does not reach the cytoplasm, the information will be lost and dramatic disturbances may occur within the cell. This transport is, however, no one-way street. It is likewise important that certain substances and proteins reach the nucleus so as to maintain the function of the cell. If a transcription factor, for example, which - like the other proteins - is formed in the cytoplasm does not reach the cell nucleus, it cannot trigger the transcription of the other genes. Dramatic disturbances of the events in the cell, which may even comprise the dying of the cell or the organism, are often accompanied by this. This shows clearly that nuclear pore proteins perform an extremely important function within the cell. The analysis of the T gene has now shown that the T protein is also incorporated into the nuclear membrane. It is interesting that the T protein is almost twice as large as the POM121 protein, i.e. it has a much greater binding capacity than the POM121 protein. The T protein is therefore very well suited to isolate possible binding partners which attach to the T protein, in particular to the C-terminus of the T protein.

The tissue-specific expression of the T gene shows strikingly that nuclear core proteins (in particular nuclear pore membrane proteins) do not have to be expressed in all cells and at all times like 'housekeeping' genes. The predominant expression of the T gene in the nervous system shows that the T protein in the nervous system performs a very specific function. The predominant expression of the T gene in the nervous system can now be used for the development of new medicaments and new medicament compound classes. New substances can now be isolated by means of the T protein, which influence deliberately the bi-directional transport in nuclear pores of the nervous system. The localization of the T protein within the nuclear membrane is in this case of major advantage. Chemical compounds can be tested by means of automated tests. Many pharmaceutical companies have suitable screening methods in which more than 200,000 chemicals can be tested. For this purpose, e.g. reporter assays (e.g. GFP fusion proteins, colored substances, etc.) can be used which show the successful transport of a molecule into the nucleus or into the cytoplasm. By this, new active substances can then be isolated which deliberately influence the transport of molecules into nuclear pores, in particular those of the nervous system.

Identifying and analyzing interactions between the T proteins according to the invention (T, T2, T3 protein) or peptides or fragments thereof and possible binding partners which may represent active substances within the above-mentioned meaning, can happen e.g. with the "yeast-two-hybrid system" (Fields, Nature 340, pages 245-247, (1989)). This system is based on the discovery that cellular transcription activators, such as GAL4 or lexA from yeast,

can be separated into two independent functional domains. Both domains are usually part of a protein in the cell nucleus of the yeast cell, which binds to certain activating sequences of different target genes and regulates the transcription thereof. In this connection, one domain, the DNA binding domain (BD), binds specifically to a certain DNA target sequence (upstream activating sequence) in the vicinity of the target promoter. The other domain, the activation domain (AD), increases the transcription rate of the target gene by interaction with the transcription initiation complex which is bound to the promoter of the target gene. In the "yeast-two-hybrid system", this structure is used by the transcription factors in modified form. The DNA binding domain (BD) of GAL4 or lexA is expressed there as fusion protein with a "bait protein or peptide" (here: T, T2 or T3 protein/peptide) in yeast cells. This fusion also has a nuclear localization signal by which it is transported into the cell nucleus of the yeast. The bait fusion protein binds therein to a target sequence (UAS) which is located in the employed yeast strain in the vicinity of the promoters of two reporter genes (e.g. auxotrophic marker (HIS3) and enzymatic marker (lacZ)). By this a constellation results in which the bait protein or peptide is exposed in direct spatial vicinity of the reporter gene promoter. Then, a second fusion protein is additionally expressed in the same yeast cell. It consists of the activation domain (AD) of GAL4 or lexA and a prey protein or peptide. It also has a nuclear localization signal. The prey fusion protein is thus also transported into the cell nucleus of the yeast. If the prey protein and the bait protein exposed on the UAS physically interact with each other, it becomes more likely statistically that the activation domain is located in the vicinity of the reporter gene promoter. This results in an increase of the



transcription of the reporter genes whose extent is proportional to the strength of interaction between bait and prey protein. In this case, e.g. a cDNA library and also a combinatorial peptide library are in consideration as the prey proteins.

The present invention also relates to a process of identifying inhibitors or enhancers of the T protein family according to the invention. For this purpose, the nucleic acid sequences or parts of these sequences, which are part of the T gene or the paralogs or orthologs thereof, are inserted in suitable vectors and used for transfecting or transforming cells, tissues or organisms. These changed cells, tissues or organisms are then used for identifying inhibitors or enhancers of the T protein or its paralog or ortholog proteins (e.g. T2 and T3) or proteins which interact directly or indirectly with these proteins. The inhibitors or enhancers identified by this approach can be used for pharmaceutical active substances or medicaments or for the production thereof and for the treatment of diseases such as cancer, neurological and psychiatric diseases and injuries of the nervous system. In the case of injuries of the nervous system, innate damage of the nervous system or the degenerative diseases of the nervous system, it is possible to support deliberately by this treatment *inter alia* the neuronal regeneration or improve the interconnection of individual nervous regions (used for *inter alia* Alzheimer's disease, Parkinson's disease, schizophrenia, manic-depressive diseases, autism, mental retardation). The present invention provides the possibility of testing the substances or therapeutic agents suitable to enhance or reduce the effect of the T protein or the family of the T proteins. In particular, the changed nuclear pore properties which are influenced by the proteins T and T3 can

be detected by suitable screening methods. The latter include e.g. visualization of the bi-directional transport through the nuclear pore or the detection of a modified transcription of cellular or reporter genes. Substances or therapeutic agents can also be identified which inhibit or promote the effect of proteins which are directly or indirectly involved in the effect of the T protein or the family of the T proteins. Substances or therapeutic agents which show an enhancement or reduction of the effect of the T protein (or T2 or T3) in the above-mentioned screening methods, can be used to determine whether the enhancement or the reduction of the effect of the T protein results in therapeutically desired effects. Above all the inhibition of the growth or spreading of tumor cells or the support of neuronal regeneration, e.g. after injuries of the nerves (*inter alia* paraplegia and head-brain trauma), are counted thereamong. The identified substances can then be used as medicaments or for the production of these medicaments. Due to these medicaments it is then possible to inhibit or block spreading of the disease-inducing cells and thus control or clear up the disease on the whole. An important application of these medicaments is *inter alia* preventing the growth and spreading of tumor cells. In addition thereto, the identified active substances are used as medicaments which stimulate deliberately the growth of certain cells. By this it is then possible to regenerate cells or structures of the nervous system damaged by injury or degenerative processes. The T protein (or T2 or T3) can also be used in screening methods allowing not only to detect the changed nuclear pore properties but also to identify prior or subsequent or parallel signal cascades. By this it is possible to identify e.g. tyrosine kinases or tyrosine phosphatases which regulate proteins which in turn influence directly or indirectly the action of the T protein (or T2 or T3). As a

result, suitable targets for the positive influence of the events in the cells can be recognized and characterized. Furthermore, the T protein, although it occurs as a nuclear pore protein, is significant for the interactions with filaments of the cell, e.g. microtubuli and actin. These interactions can now be studied, e.g. by fusion proteins of the T protein with the EGFP protein. Cells which were stably or transiently transformed or transfected with constructs for such fusion-reporter proteins, can be incubated with substances or pharmaceutical preparations to identify substances which enhance or reduce the interaction of the T protein with filaments such as the actin filaments or the microtubuli. As a result, it is possible to isolate active substances which positively influence *inter alia* the growth of nerve cells or the inhibition of the growth of tumor cells. For example, immunoprecipitation has to be mentioned as a method of identifying such possible active substances. Proteins can be isolated by this which bind to the T protein family. Further immunoprecipitations can then be carried out with these proteins to isolate new proteins which then no longer interact directly with the T protein.

The present invention also relates to a method of identifying further proteins which play a role in the development and function of the nervous system and/or are a nuclear pore protein, the method comprising the steps of:

- (a) producing an antibody against a protein of the T family (T, T2 or T3 protein),
- (b) contacting a cell extract with the antibody and identifying the antibody/protein complex,

- (c) analyzing the complex to identify a protein which has bound to the protein of the complex and is no antibody, and
- (d) optionally repeating steps (a) to (c) to identify further proteins of this function.

The invention is described in more detail by means of the figures, which show:

Figure 1: human cDNA sequence (gene T) and derived amino acid sequence

Figure 2: human genomic DNA sequence (gene T)

Figure 3: human genomic DNA sequence (gene T)

Figure 4: human genomic DNA sequence (gene T)

Figure 5: human genomic DNA sequence (gene T)

Figure 6: human genomic DNA sequence (gene T)

Figure 7: human genomic DNA sequence (gene T)

Figure 8: human genomic DNA sequence (gene T)

Figure 9: partial murine cDNA sequence (gene T) and derived amino acid sequence

Figure 10: partial murine genomic DNA sequence (gene T)

Figure 11: partial human cDNA sequence (gene T2) and derived amino acid sequence

Figure 12: partial murine cDNA sequence (gene T2) and derived amino acid sequence

Figure 13: partial murine cDNA sequence (gene T2) and derived amino acid sequence

Figure 14: splicing variant of the human T gene with derived amino acid sequence

Figure 15: splicing variant of the human T gene with derived amino acid sequence

Figure 16: partial human cDNA sequence (gene T2) with  
derived amino acid sequence

Figure 17: partial human cDNA sequence (gene T3; protein isoform 1) with derived amino acid sequence

Figure 18: partial human cDNA sequence (gene T3; protein isoform 2) with derived amino acid sequence

Figure 19: partial murine cDNA sequence with derived amino acid sequence (gene T3)

Figure 20: oligonucleotide and peptides (T gene)

Figure 21: sequence comparison within the T family

Figure 22: protein alignment of POM121 protein and T protein

Figure 23: Northern blot analysis

Figure 24: immunohistochemical studies and electron-microscopic pictures

Figure 25: *in situ* hybridization with embryonal RNA

Figure 26: *in situ* hybridization with RNA from brain

Figure 27: *in situ* hybridization with RNA from fetal brain

Figure 28: *in situ* hybridization with RNA from nerve tissues of mouse

Figure 29: comparison of the coiled-coil regions between CLIP protein, T protein and POM121

Figure 30: hydrophobicity blot for POM121, T protein and T3 protein.

The following clones were deposited with the DSMZ (*Deutsche Sammlung für Mikroorganismen and Zellkulturen GmbH*) [German-type collection of microorganisms and cell cultures], Mascheroder Weg 1b, Braunschweig, according to the Budapest treaty on August 18, 1998:

- clone JFC277 (DSM12371); human cDNA; represents the human cDNA sequence of Bp 1218-3690
- clone JFC405 (DSM12372); human cDNA; represents the human cDNA sequence of Bp 1-1891
- clone JFC601 (DSM12373); murine cDNA; represents the murine cDNA sequence of Bp 225-3026
- clone JFC950 (DSM12374); human genomic clone; represents human genomic sequence

- clone JFC955 (DSM12375); human genomic clone; represents human genomic sequence; comprises start of the cDNA sequence
- clone JFC N2112 (DSM12376); human genomic clone; was fully sequenced. The sequence is shown in figure 2 and contains the sequence of Bp 1756-4228 of the human cDNA sequence.

The following clone was deposited with DSMZ according to the Budapest treaty on February 2, 1999:

- clone JFC-BN27 (DSM 12659); contains the sequence of Bp 4370-8690 of the human cDNA sequence.

The following clone was deposited with the DSMZ according to the Budapest treaty on February 19, 1999:

- clone JFC-BN20 (DSM 12698); contains the sequence of Bp 2025-6280 of the human cDNA sequence

The following clone was deposited with the DSMZ according to the Budapest treaty on February 1, 2000.

- cDNA clone pL70 (DSM13270); represents essential parts of the gene T3.

The sequences shown in figures 2 to 8 originates from clones JFC955 (DSM 12375) and JFC950 (DSM 12374). The sequence shown in figure 1 originates from clones JFC277 (DSM 12371), JFC405 (DSM 12372) and JFC-BN27 (DSM 12659) and JFC-BN20 (DSM 12698). The sequence shown in figure 9 originates from the clone JFC610 (DSM12373).

The invention is further described by means of the following embodiment.

### EXAMPLES

As to the methods employed reference is also made to Sambrook, J., Fritsch, E.F. and Maniatis, T. (Molecular Cloning; A Laboratory Manual; second edition; Cold Spring Harbor Laboratory Press, 1989) and Current Protocols in Molecular Biology (John Wiley and Sons, 1994-1998), the below techniques, in particular preparation of DNA or RNA or Northern blot, being sufficiently known to, and mastered by, the person skilled in the art.

Before it is described in detail how the experiments are carried out, the operating strategy is to be explained first.

When screening for genes triggering diseases of the CNS (e.g. neurodegenerative diseases, mental retardations, tumoral diseases of the CNS) in the mutated state, 23 cDNA clones were isolated from a human fetal brain cDNA library (Stratagene company, Heidelberg). A human fetal brain cDNA library was used as a starting material, since it was assumed that genes which play a role in the development of the CNS and in particular of the brain are present in a fetal brain cDNA library. However, since what is called housekeeping genes (genes expressed in most tissues) are also expressed in the CNS, it was tested whether the selected cDNA clones originate from genes having a CNS-specific expression. For this purpose, the cDNA pieces ('inserts') contained in the individual cDNA clones were isolated and used for hybridization with Northern blots. The employed



### EXAMPLE 1: Identification of the T genes

In order to ensure an effective infection, it was initially necessary to produce phage-competent bacteria in an overnight culture. The magnesium ions contained in the medium induce the maltose receptor of the bacteria to which the phage binds to infect the bacterium.

Charge 50  $\mu$ l *E. coli* XL1-Blue in 50 ml LB broth, the medium being admixed with  $MgSO_4$  in a concentration of 10 mM. Incubate overnight at 30°C and 220 rpm. Centrifuge off the bacteria at 4°C and 1000 xg for 10 min. Resuspend in 25 ml 10 mM  $MgSO_4$ . The thus produced phage-component bacteria could be stored at 4°C for up to one week.

## 2. Culturing the cDNA libraries

For culturing the library, Baltimore Biological Lab. (BBL) agar plates and BBL top agarose had to be prepared. The phages (human or murine cDNA library, Stratagene company) were mixed with SM medium to a dilution of  $1:10^3$  and  $1:10^4$  to obtain individual plaques after the culturing.

#### Performance:

For the BBL agar (pH 7.2) 10 g BBL trypticase, 5 g NaCl and 10 g Select agar were weighed and filled to 1 l with H<sub>2</sub>O. The agar is dissolved by autoclaving. After cooling to about 60° pour the plates. The plates are preheated to 37°C prior to their use to avoid premature solidification of the top agarose. The BBL top agarose (pH 7.2) was prepared with 10 g BBL trypticase, 5 g NaCl, 6.5 g agarose and 10 ml 1 M MgSO<sub>4</sub> solution. Dissolve by autoclaving and provide in the water bath to 41°C. Add 15 µl of the above indicated dilute phage solution and 250 µl of the competent XL-1 bacteria in a 15 ml Falcon tube. Incubate at room temperature for 20 minutes. Add 10 ml BBL top agarose, swivel and place on the heated agar plate. The top agarose layer is solid after about 20 minutes and the plates can be stacked with the agar side up. Incubation is carried out overnight at 37°C. The plates can be stored at 4°C after expired incubation time or can be used directly for transferring the phage plaques. Carefully close the plates for storing them together with a chloroform-soaked cloth in plastic bags. The chloroform prevents the growth of cryophilic bacteria and fungi.

### 3. In vivo excision

The employed cDNA libraries (human and murine fetal brain cDNA library; Stratagene company, Heidelberg) were cloned in the vector λ-ZAPII. Due to this there was the possibility of circumventing the subcloning of the phage insert in a plasmid vector. This protocol permits to transfer cDNA which

is located as insert in the  $\lambda$ -ZAPII vector into an insert in simple way by an *in vivo* preparation which is now found in the plasmid Bluescript SK(-). In principle, this preparation serves for introducing by a helper phage information for proteins which permit DNA amplification only in the region of the phage genome, which have the genetic information for the plasmid with cDNA insert. For the most part, the method was carried out in accordance with the protocol of the manufacturer (Stratagene).

In particular, culturing was made such that individual phage plaques were on the plate. Then, the *in vivo* excision protocol was carried out with these individual plaques. The plasmid DNA and its plasmid inserts were isolated from the bacterial clones and subsequently hybridized with Northern blots. The selection of further clones to be studied was based on the expression pattern in the Northern blots.

#### Performance:

Mix 100  $\mu$ l of a single phage  $\lambda$ -ZAPII clone with 200  $\mu$ l XL1 bacteria and 2  $\mu$ l helper phages (contained in the Stratagene kit). Shake for 15 min. at 37°C and 80 rpm, the specific attachment of both phage types to the host bacterium taking place. Add 3 ml LB broth. Incubate for 2 h at 37°C and 200 rpm. The DNA replication of the plasmid contained in the  $\lambda$ -ZAPII, its circularization and the packing into coat proteins take place and discharge from the bacterium occur during this time. Heat to 70°C for 20 minutes. Thereafter, centrifuge at 4000 g for 15 minutes. This kills the still remaining bacteria and separates their fragments from the plasmids existing in the phage coat, which are found in the supernatant. Add 1  $\mu$ l thereof to 200  $\mu$ l SOLR host cells, incubate at 37°C for 15 minutes. Plate 100  $\mu$ l onto LB/Amp plates. Store at 37°C overnight. The then grown bacterial

clones contain the plasmid with the corresponding cDNA insert. A mini-prep DNA preparation was carried out each.

#### **4. "random primed" DNA labeling**

The radioactive labeling of the double-stranded insert DNA of the cDNA clone was carried out as follows for the further isolation of overlapping cDNA clones:

##### **Performance:**

Dissolve 100 ng DNA in a volume of 12  $\mu$ l H<sub>2</sub>O for a typical labeling batch. 10-minute heating to 95°C effects the denaturation of the DNA into single strands. Store the preparation on ice to prevent reassociation of the two complementary DNA strands. Complete the reaction batch by 4  $\mu$ l OLG (oligo-labelling buffer), 1  $\mu$ l Klenow (1U) and 2.5  $\mu$ l  $\alpha$ -<sup>32</sup>P-dCTP and 2.5  $\mu$ l  $\alpha$ -<sup>32</sup>P-dATP. Incubate at room temperature overnight. Based on the hexanucleotides attached to a single strand, the formation of the complementary strand takes place during this time by the Klenow fragment of the *E. coli* DNA polymerase I. The DNA is labeled radioactively by incorporating  $\alpha$ -<sup>32</sup>P-dCTP and the  $\alpha$ -<sup>32</sup>P-dATP.

#### **5. Separation of non-incorporated radioactive nucleotides**

The non-incorporated nucleotides were separated by means of a personally prepared sephadex G-50 column. The separation principle of the column is based on the exclusion chromatography. The smaller non-incorporated nucleotides fit into small pores of the column material while the DNA is locked out. The volume in which the nucleotides may move is thus greater than the volume available to the DNA. If a mixture of DNA and nucleotides is placed on the column, the DNA runs through the column faster than the nucleotides. This permits the separation of non-incorporated nucleotides.

Performance:

A Pasteur pipette was closed with a small glass bead. Fill the Pasteur pipette with sephadex G-50 ("fine") dissolved in water until the filling material is 5 cm below the top edge of the Pasteur pipette. Rinse the column 2 times with TE. Apply the above radioactive labeling batch. Add 320  $\mu$ l TE. Discard the solution which has run through the column. Place an Eppendorf tube below the column. Add 350  $\mu$ l TE. Collect the radioactive solution run through the column.

## 6. Plaque "blot"

The plaque "blot" was made to analyze the cDNA library to make accessible the cDNA in the phage clones to hybridization.

Performance:

Place a labeled hybond-N membrane provided with an inscription in air bubble-free manner on the plate with the phage plaques for one minute. The labeling pattern was transferred. Place it on a Whatman paper soaked with denaturing solution (0.5 M NaOH; 1.5 M NaCl) for 10 minutes. Neutralize in 50 mM phosphate buffer for 10 minutes. The rests of the bacterial layer are wiped off with slight pressure using a phosphate buffer-soaked Kleenex cloth. The filters are spread at room temperature for drying. Thereafter, the filters were baked at 90°C for 1 h.

## 7. Hybridization

The hybridization is based on the binding of complementary, single-stranded nucleic acids. For this purpose, the DNA to be studied was immobilized on a membrane and hybridized with a radioactively labeled probe. The complementary binding is maintained even after washing off the non-specifically adhering probes and can be made visible by means of

autoradiography. Single-stranded molecules were incubated during the hybridization under salt and temperature conditions which support the formation of base-paired double strands. A decisive factor in the association and dissociation kinetics are the hydrogen bridge bonds between the base pairs G-C and A-T. The hybridization reaction is influenced by changing the temperature and the salt and sample concentrations.

#### Performance:

First, prehybridize the DNA filters in hybridization solution (0.5 M NaPi (pH 7.2); 7 % SDS; 0.2 % BSA; 0.2 % PETG 6000; 0.05 % polyvinyl pyrrolidone 360000; 0.05 % Ficoll 70000; 0.5 % dextrane sulfate) with a 0.1 ml/cm<sup>2</sup> at 65°C. For this purpose, incubate the filters in a plastics box in a shaking water bath at 65°C for a period of at least 1 h. Discard the prehybridization solution. Place the radioactively labeled sample (see above items 4. and 5.) with 0.5 ml/cm<sup>2</sup> of hybridization solution (65°C) on the filters. The activity of the sample should not drop below 50 cpm, measured at a distance of 40 cm. The hybridization takes place overnight at 65°C (human cDNA library) or 55°C (interspecies hybridizations man-mouse and for isolating the homologous genes). Wash the filters two times for 30 minutes with about 500 ml wash buffer in a shaking bath at 65°C (55°C). Autoradiography was then carried out.

### **8. Autoradiography**

The filters were packed in plastic foodwrap. The autoradiography was made at -80°C in an X-ray cassette containing a reinforcing film made of calcium tungstate. The exposure is 30 minutes to several days, depending on the strength of the signal.

The complete mRNA which codes for the protein of the T gene could be isolated by means of the above mentioned techniques. Furthermore, using cDNA clones of this newly isolated T gene it was possible to isolate two further genes (T2 and T3) which have distinct homologies with this gene. For this purpose, the above mentioned techniques were used again. For isolating the related genes T2 and T3, the hybridization temperature was lowered to 55°C.

#### **EXAMPLE 2: Northern blot**

The 'multiple tissue Northern blots' were purchased from the CLONTECH company (Palo Alto, California, U.S.A.) and used in accordance with the instructions from the manufacturer. The respective DNA samples of the genes T, T2 and T3 were labeled radioactively and hybridized with the Northern blots. The sequence of bp 1-4200 of figure 1 was used for the analysis of the expression pattern on a Northern blot level. For the gene T3 the sequence of bp 1310-4870 of figure 17 was used for hybridization. The sequence of bp 3120-4230 of figure 16 was used for the gene T2. The "random priming" method was used for the radioactive labeling of double-stranded DNA.

##### a) Random priming:

Dissolve 100 ng DNA in a volume of 12 µl for a typical labeling batch. 10-minute heating to 95°C effects the denaturation of the DNA into single strands. Store the batch on ice to prevent reassociation of the two complementary DNA strands. Complete the reaction batch by 4 µl OLB, 1 µl Klenow (1U) and 2.5 µl a-<sup>32</sup>P-dCTP and 2.5 µl a-<sup>32</sup>P-dATP. Incubate at room temperature overnight. Based on the hexanucleotides attached to a single strand, the formation of the complementary strands takes place during this time by

The non-incorporated nucleotides were separated by means of a personally prepared sephadex G-50 column. The separation principle of the column is based on the exclusion chromatography. The smaller non-incorporated nucleotides fit into small pores of the column material while the DNA is locked out. The volume in which the nucleotide may move is thus greater than the volume available to the DNA. If a mixture of DNA and nucleotides is placed on the column, the DNA runs through the column faster than the nucleotides. This permits the separation of non-incorporated nucleotides. For this purpose, a Pasteur pipette is closed with a small glass bead. Fill the Pasteur pipette with sephadex G-50 ("fine") dissolved in water until the filling material is 5 cm below the top edge of the Pasteur pipette. Rinse the column 2 times with TE. Apply the above radioactive labeling batch. Add 320  $\mu$ l TE. Discard the solution which has run through the column. Place Eppendorf tube below the column. Add 350  $\mu$ l TE. Collect the radioactive solution run through the column.

The Northern blots were hybridized as described below. First, the Northern blots were prehybridized at 65°C in 10 ml hybridization solution (350 ml 20 % SDS, 500 ml 1 M phosphate buffer, pH 7.2; 150 ml distilled water). For this purpose, the Northern blots were incubated in a glass tube in a hybridization roll-over-type furnace at 65°C for a period of 6 h.



The prehybridization solution was discarded. The radioactively labeled sample was placed with 10 ml hybridization solution (65°C) on the filters.

The hybridization was carried out at 65°C overnight. The filters were then washed two times for 30 min. with about 500 ml wash buffer (80 ml 1 M phosphate buffer, pH 7.2; 100 ml 20 % SDS, 1820 ml distilled water) at 65°C in a shaking bath.

### c) Autoradiography

The filters were welded into plastic film. The autoradiography was made at -80°C in an X-ray cassette which contained a reinforcing film of calcium tungstate. Exposure was 1 to 4 days depending on the strength of the signal.

The results of the Northern blots carried out are shown in figure 23.

### **EXAMPLE 3: RNA in situ hybridization**

Embryos in various development stages were isolated from pregnant NMRI mice. The embryos and other tissue samples were fixed overnight with 4 % paraformaldehyde in PBS at 4°C. 10 µm freezing sections of the embryos were transferred to slides coated with 3-aminopropyl triethoxysilane. Sense strand ("sense") and antisense strand ("antisense") samples were produced by transcription with  $\alpha$ -<sup>35</sup>S-UTP with a specific activity of  $>10^9$  decays per minute/µg. For this purpose, the linearized mouse T gene cDNA clone from figure 9 was transcribed with T7 or Sp6-RNA polymerase. The sample length was reduced by alkaline lysis to 150 to 200 nucleotides. The slides were prehybridized at 54°C in a solution containing 50 % formamide, 10 % dextrane sulfate, 0.3 M NaCl, 10 mM Tris, 10 mM sodium phosphate, pH 6.8, 20

mM dithiothreitol, 0.2 % Denhardt's solution, 0.1 Triton X-100, 0.1 mg/ml Escherichia coli RNA and 0.1 mM non-radioactive  $\alpha$ -S-UTP. The  $^{35}\text{S}$ -labeled sample ( $8 \times 10^4$  decays per minute per ml) were added to the hybridizing mixture for the hybridization and the hybridization was then continued for 16 h at  $54^\circ\text{C}$  in a humid chamber. The slides were then washed in the hybridization solution for 2 hours. The remaining non-hybridized RNA sample was then digested using RNase A. Thereafter, the slides were washed for 30 minutes at  $37^\circ\text{C}$  with 2x SSC, 0.1 % SDS and for 30 minutes with 0.1x SSC, 0.1 % SDS. Then, the slides were dehydrated with increasing ethanol concentrations. The slides were covered with Ilford K5 autoradiography emulsion. After 1 to 2 weeks of exposure at  $4^\circ\text{C}$ , the slides were incubated in Kodak D19b developer and dyed with Giemsa. The sections were analyzed in dark field and bright field illumination with a Zeiss SV8 stereomicroscope and an Axiophot microscope and photographed with an Agfa ortho black-and-white film.

The results of the RNA *in situ* hybridization are shown in figures 25, 26, 27 and 28.

Figure 25: expression of the murine T gene during the mouse embryogenesis. Bright field (a,c,e,g) and dark field pictures (b, d, f, h) of horizontal (a,b) and sagittal sections (c-h) through a 10.5 (a,b), 12.5 (c,d), 14.5 (e,f) and 16.5 (g,h) dpc embryo (dpc = days post conceptionem) which were hybridized with an antisense ribo sample of the murine T gene. Dec = decidua, g = guts, he = heart, lab = labyrinth, li = liver, me = myelcephalon, sc = spinal cord, sga = spinal ganglia, sb = tooth bud, te = telencephalon. Bar = 1 mm.

Figure 26: Expression of the murine T gene in the postnatal brain. Bright field (a,d) and dark field pictures (b,c,e,f) of horizontal sections through an 1 wpn (weeks *post natalis*) and 6 wpn head, which were hybridized with a T gene antisense (b,e) and a sense sample (c,f). cd = cerebellum, cor = cortex, cos = colliculus, ey = eye, hi = hippocampus, ne = nasal epithelium, ob = olfactory bulb, bar = 1 mm.

Figure 27: Greater enlargement of the 10.5 dcp embryo of figure 25 a,b. The arrows point to a region of little expression in the somites (arrows in b). An intense expression can be seen in the region between mesencephalon and telencephalon ("forebrain-midbrain junction"). Aod = aorta dorsalis, me = mesencephalon, sc = spinal cord, te = telencephalon. Bar = 100  $\mu$ m.

Figure 28: Expression of the T gene during the development of the nervous system. Expression of the T gene in neurons of the mantle zone of the developing brain and in nuclei of peripheral nerves (arrow in b). No expression is visible in proliferating neurons in the subventricular layer or in migrating neurons of the intermediate zone (c,d). On day 16.5, an intense expression is visible in differentiating neurons of the mantle zone of the telencephalon (e,d). A minor expression is also visible in neurons of the spinal cord and the spinal ganglia (g,h). Furthermore, a minor expression is visible in an individual layer below the skin (g,h). iz = intermediate zone, mz = mantle zone, sc = spinal cord, sga = spinal ganglia, sk = skin, svl = subventricular layer, vn = ventricle. Bar = 100  $\mu$ m.

#### EXAMPLE 4: Production of antibodies

Using a synthetically produced peptide of the sequence "EKGEDPETRRMRTVKNIAD" animals are immunized to produce antibodies against the T protein as follows:

Immunization protocol for polyclonal antibodies in rabbits

600 µg purified KLH-linked peptide in 0.7 ml PBS and 0.7 complete or incomplete Freund's adjuvant are used per immunization:

- Day 0: 1<sup>st</sup> immunization (complete Freund's adjuvant)
- Day 14: 2<sup>nd</sup> immunization (incomplete Freund's adjuvant; icFA)
- Day 28: 3<sup>rd</sup> immunization (icFA)
- Day 56: 4<sup>th</sup> immunization (icFA)
- Day 80: bleeding to death.

The rabbit serum is tested in an immunoblot. For this purpose, the protein used for the immunization is subjected to SDS polyacrylamide gel electrophoresis and transferred to a nitrocellulose filter (*cf.* Khyse-Andersen, J., J. Biochem. Biophys. Meth. 10 (1984), 203-209). The Western blot analysis was carried out as described in Bock, C.-T. et al., Virus Genes 8, (1994), 215-229. For this purpose, the nitrocellulose filter is incubated with a first antibody at 37°C for one hour. This antibody is the rabbit serum (1:10000 in PBS). After several wash steps using PBS, the nitrocellulose filter is incubated with a second antibody. This antibody is an alkaline phosphatase-coupled monoclonal goat anti-rabbit IgG antibody (Dianova company) (1:5000) in PBS. 30 minutes of incubation at 37°C are followed by several wash steps using PBS and subsequently by the alkaline phosphatase detection reaction with developer solution (36 µM 5'-bromo-4-chloro-3-indolylphosphate, 400 µM nitro blue tetrazolium, 100 mM Tris-HCl, pH 9.5, 100 mM

It shows that polyclonal antibodies according to the invention can be prepared.

100 µg of purified KLH-linked peptide in 0.8 ml PBS and 0.8 ml of complete or incomplete Freund's adjuvant are used per immunization.

Antibodies are extracted from egg yolk and tested in a Western blot. Polyclonal antibodies according to the invention are detected.

Day 0: 1<sup>st</sup> immunization (complete Freund's adjuvant)  
Day 28: 2<sup>nd</sup> immunization (incomplete Freund's adjuvant;  
icFA)  
Day 56: 3<sup>rd</sup> immunization (icFA)  
Day 84: 4<sup>th</sup> immunization (PBS)  
Day 87: fusion.

### EXAMPLE 5: immunohistochemical studies

Prepare an 1:10 dilution of normal (sheep) serum in PBS (e.g. sheep Dako X0503, Dako company, Hamburg), add 100  $\mu$ l thereof and incubate for 20 minutes.

Rinse slides in PBS, pour off, rinse again, thereafter allow to stand in PBS for 10 minutes.

Take out slide and wipe off the liquid around the tissue using a cloth.

Add first antibody in a dilution of 1:100.

Add 100  $\mu$ l of the first antibody (in PBS) and incubate in a refrigerator in a humid chamber overnight. Control: without first antibody.

2<sup>nd</sup> day

Take humid chamber out of the refrigerator and allow to stand at room temperature. Rinse slide in PBS, pour off, rinse again, thereafter allow to stand in PBS for 10 minutes, when many slides are analyzed wash two times with PBS.

Take out slides and wipe off the liquid around the tissue using a cloth.

Prepare a 1:100 dilution of second antibody, "antirabbit biotinylated" (Amersham company, Braunschweig) in PBS and add 100  $\mu$ l thereof.

Incubate in a humid chamber at room temperature for 45 minutes.

Rinse slides in PBS, pour off, rinse again, thereafter allow to stand in PBS for 10 minutes.

Take out slide and wipe off the liquid around the tissue using a cloth.

Prepare a 1:100 dilution of streptavidine peroxidase (streptavidine horseradish) (Amersham company, Braunschweig) with PBS and add 100  $\mu$ l thereof.

Incubate in a humid chamber at room temperature for 45 minutes.

Rinse slides in PBS, pour off, rinse again, thereafter allow to stand in PBS for 10 minutes.

Take out slides and wipe off the liquid around the tissue using a cloth.

Staining: Add one drop chromogen per ml buffer just before the use. Vortex and place in the dark.

Add 100 µl staining solution (Dako company, Hamburg).

Finally, stain the control. Incubate for about 2 minutes.

Incubate slides in water. Inspect under a microscope.

Place 1-2 drops of crystal Mount on the section. If there is an air bubble, suck it off with a paper handkerchief.

The rest of the slide is wiped doff using HCl-EtOH to remove the stain.

Place a line of adhesive (Eukitt) on the cover glass. Press the cover glass onto the slide without producing air bubbles.

The enzyme in the second antibody results in a dye formation (DAB) so that the T protein can be detected.

Figure 24 (a-d): Light-microscopic pictures which show that the T protein is localized in or at the nucleus of the cell. The electron-microscopic picture in e shows that the T protein is not localized in the nucleus but in the membrane. The pictures are highly consistent with a function as a membrane-terminal nuclear pore protein. The arrows in e show the stain formed which can be seen on the cytoplasmic side of the nuclear membrane.

...





3. An antisense RNA, characterized in that it is complementary to the DNA sequence of claim 1 or 2 and can reduce or inhibit the synthesis of the protein encoded by this DNA sequence.
4. Ribozyme, characterized in that it is complementary to the DNA sequence of claim 1 or 2 and can bind specifically to the RNA transcribed by this DNA sequence and can cleave it so as to reduce or inhibit the synthesis of the protein encoded by this DNA sequence.
5. Expression vector, containing the DNA sequence according to claim 1 or 2 or coding for the antisense RNA according to claim 3 or the ribozyme according to claim 4.
6. The expression vector according to claim 5, which comprises additionally the promoter of the human T gene or an ortholog of the T gene.
7. Expression vector according to claim 5 or 6, which codes for the T, T2 or T3 proteins or for fragments thereof in the form of a reporter fusion protein.
8. Host cell which is transformed with the expression vector according to any of claims 5 to 7.
9. Protein which is encoded by the DNA sequence according to claim 1 or 2 and which is involved in the development of the nervous system and is expressed in tissue-specific and development-specific manner, or fusion proteins, fragments, variants, derivatives or precursors of the protein.



14. Use of the DNA sequence according to claim 1 or 2, the antisense RNA according to claim 3, the ribozyme according to claim 4, the expression vector according to any of claims 5 to 7, the protein according to claim 9 or the antibody or the fragment thereof according to claim 12 or 13 for preventing or treating diseases of the nervous system, in particular of the CNS.
15. Use according to claim 14, wherein the disease of the nervous system is a tumoral disease of the CNS.
16. Use according to claim 14, wherein the treatment of diseases of the nervous system are the promotion of the neuronal regeneration in the case of injuries of the nervous system and degenerative diseases of the nervous system.
17. Use according to claim 14, wherein the treatment of diseases of the nervous system are the regeneration of the neuronal linkages and the regeneration of the innate and acquired malfunctions of the nervous system.
18. Use according to claim 15 for inhibiting the growth and spreading of tumor cells.
19. Diagnostic method for detecting a disturbed expression of the protein according to claim 9 or for detecting a changed form of this protein, in which a sample is contacted with the DNA sequence according to claim 1 or 2 or the antibody or the fragment thereof according to claim 12 or 13 and then it is determined directly or indirectly whether the concentration of the protein

and/or its amino acid sequence differs from a protein obtained from a healthy patient.

20. Diagnostic kit for carrying out the method according to claim 19, which contains the DNA sequence according to claim 1 or 2 and/or the antibody or the fragment thereof according to claim 12 or 13.
21. Non-human mammal whose naturally occurring T, T2 or T3 gene comprises a change in the gene structure or the gene sequence.
22. Non-human mammal, wherein a change of the gene structure of the T, T2 or G3 gene is achieved in the mammal by introducing a deletion in place of which a homologous or heterologous sequence is introduced.
23. Non-human mammal, wherein a change of the gene structure of the T, T2 or G3 gene is achieved by inserting a homologous or heterologous sequence in the corresponding gene naturally occurring in the mammal.
24. Non-human mammal according to claim 22 or 23, wherein the heterologous sequence is the selection marker sequence.
25. Non-human mammal according to claim 24, wherein the selection marker sequence conveys resistance to neomycin.
26. A method of producing a non-human mammal according to any of claims 21 to 25, characterized by the steps of:
  - (a) producing a DNA fragment, in particular a vector, containing a changed T, T2 or G3 gene, the T, T2

or T3 gene having been modified by inserting a heterologous sequence, in particular a selectable marker;

- (b) preparing embryonal stem cells from a non-human mammal (preferably a mouse);
- (c) transforming the embryonal stem cells from step (b) with the DNA fragment from step (a), the T gene in the embryonal stem cells being changed by homologous recombination with the DNA fragment from (a)
- (d) culturing the cells from step (c),
- (e) selecting the cultured cells from step (d) for the presence of the heterologous sequence, in particular the selectable marker,
- (f) producing chimeric non-human mammals from the cells of step (e) by injecting these cells into mammalian blastocysts (preferably mouse blastocysts), transferring the blastocysts to pseudo-pregnant female mammals (preferably mouse) and analyzing the resulting offspring for a change of the T, T2 or T3 gene.

27. Transgenic cell or tissue which is capable of expressing a T protein or part of the T protein or an ortholog thereof.

28. Use of the non-human mammal according to any of claims 21 to 25 or the transgenic cell or the transgenic tissue according to claim 27 for the analysis of the function of the T gene family.

29. Use of the non-human mammal according to any of claims 21 to 25 or the transgenic cell or the transgenic

tissue according to claim 27 for identifying inhibitors and enhancers of the T gene family.

30. Vertebrate gene and functional equivalent, derivative or a bioprecursor thereof, which code for a protein having a statistically significant amino acid sequence homology to the T gene, T2 gene or T3 gene according to any of the following figures: figure 1, figure 9, figure 11, figure 12, figure 13, figure 14, figure 15, figure 16, figure 17, figure 18 or figure 19.
31. T gene and its vertebrate orthologs and vertebrate paralogs which code for a nuclear pore protein.
32. Vertebrate protein which has an amino acid sequence according to figure 1 or an amino acid sequence which differs from the amino acid sequence in figure 1 by one or more amino acids.
33. Vertebrate T, T2 or T3 gene and the protein encoded therein in all of its naturally occurring allelic and mutated forms.
34. Medicament containing a protein according to claim 9 or a functional equivalent, a fragment or a bioprecursor thereof in combination with a pharmaceutically acceptable carrier.
35. The method of identifying substances which has an enhancing or inhibiting influence on the effect of T protein, T2 protein or T3 protein, by means of
  - determining the bi-directional transport through the nuclear pores,

- determining the binding to filaments of the cell (e.g. actin filaments and microtubuli) or
  - determining the increased or reduced transcription of cellular or reporter genes.
36. Method of identifying substances which have an enhancing or inhibiting influence on the effect of proteins which are functionally linked to the T protein in direct or indirect way, or represent parallel signal or functional pathways, by means of
- determining the bi-directional transport through the nuclear pores,
  - determining the phosphorylation and the dephosphorylation of proteins,
  - determining the binding of the T protein to filaments of the cell (e.g. actin filaments and microtubuli), or
  - determining the increased or reduced transcription of cellular or reporter genes.
37. The method according to claim 35 or 36, wherein the modified transcription with reporter molecules, preferably the occurrence of certain mRNAs or the EGFP protein, is detected.
38. The method of identifying further proteins which play a role in the development and function of the nervous system and/or are a nuclear pore protein, wherein the method comprises the steps of:
- (a) producing an antibody against a protein according to claim 9,
  - (b) contacting a cell extract with the antibody and identifying the antibody/protein complex,



- (c) analyzing the complex to identify a protein which has bound to the protein of the complex and is no antibody, and
- (d) optionally repeating steps (a) to (c) to identify further proteins of this function.

**Abstract of the Disclosure**

The invention relates to a protein (TP) and to proteins related thereto, which are involved in the development of the nervous system, especially the central nervous system, and are expressed in a tissue-specific and development-specific manner as well as to DNA sequences coding for these proteins. The invention also relates to antibodies directed against these proteins or fragments thereof and to anti-sense RNA or ribozymes which are directed against the expression of said protein. Finally the invention concerns medicaments and diagnostic processes in which the above mentioned compounds are used. The invention further relates to a non-human mammal whose TP-coding gene is modified.

1	CTCTCAGCGGCCCAAGCAGATTCTTTCTGGGTGACAGAATAATGCGCTCGGPIGGTTTTCTTTTTTCTCCCACTCTCTTAAGACGATTTCCCATAGTAACTCGATCAAGTGGCTCAAAAT
121	CGCAAAACCTCAGGATTTCCCGCGGCCCGCGCACAAGACCCTCGGCCCAGGTTTACAGGAATCTGTCTATTTTATTAATAAATGGAAACCTGTGNAAGAAAGAAATAGATAGCAGTTGAAGTCA
241	AATCTCGGATGACTATTTTGGCTTTTGGAGATCAGCATTTTAAACAGCATATGCTGATTTGGNAAGGTCCTGGGAGTAACATCGCAACCTTTATTTTCTCCATTCAAATCGATTTTTTT
361	ATCATTTCTTGGAGTCGAGTGAAGTTTCGGAACCGGTGTGATGGGGAACGTTGGCGGCCAGCTGTCTCTAGAAATATGCAATCTTGGATATAGTTTCTGCTGCTTTTTTGTGAAGAGATTCACATTT
481	TGAAGGGCAAGAACCTAATGTATGATGGATTTATCTTCAGAAATGAACAGACATGGGAAGAAATCCAGTGAGTCAAGCTAGAAAGATCAGAAAGAAATTTTACATGACCTTGGGCCCAACCACTA
601	CCTAGCAAAATCAGGCCACAAGGGCTGATCAAGGACTTTCGAACAGACATTTGCAGATGGAGTACTCTTAGCAGAAATCATCCAGATATTGCAATGAAAAGTTTGAAGATATCAATG
721	ATGCTCTAGAAAGTCTCAGATGATTGAATACTTTGATGTCTGCTCTAGTTTCTTAGCAGCGAGGGTAAATGTCTCAAGGTCTATCTGCTCAAGAAATTAAGAAATGGAACCTTAA
841	AGCCATTTCTAGGGCTGTTTTTTCAGTTTATCTCGCTACAGCACAACACCATCAACAACAGTACTATCATGCTCTTGGTGTGAACCTCAGCAGCGAGTTTACTCACGCTTCCCTTCCATC
961	GGAAGCCAGCCAGCCCAAAACCCAGCAAGATATGACAGTCCAGGCTTCCAGGGCCCTCTTAGGTCGCTCTGAGGAAGCAGCAGCAAGGTCAGGGAGCCCTCTAATTTTAAATAGGAGAA
1081	TCAGAGCTTTTAACAGCATTTGACAAAAACAAGCCCTCCAAATATGCAATGGAACAGAAAGGTGAAGACCCCTGAACAACAGAAAGANTGAGAACAGATTAAAAACATATAGCAGACTTGAAGCA
1201	GAATTTAGAAGAGACTATGTCTCAGTCTTCGTGGGACTCAGTAAAGCCACAGCACCCCTGAGACACATTTGACAGCACTGTGACAAACAGAAAGTTAATGGAAGGACCAATACCCCAACTTGAC
1321	AAAGTCGACCCCAACCCCATGACCTGAGAGGTTGGGCGCAGGCTGTCGCGACTTTCAGCGGAGGATGCTCCCTTCCCTGGGTGCTGGCTATCTCCGAGTGGTACCAGTTCATTCACACAC
1441	AGACCCCTCGAGGTTTATGATATACACGCTCTCGTCTGAGTCTCGTCTGAGTGGGAAACATGTCACAGATTTGACATGACTGAGAAAGCAAGCAGTGCACCTGGAGCATGCTTCTGA
1561	GGTCGATGTGGTGGATATATGATGTAGTGGTATATCCCTTGGGAAAGTCTCAGGACTGATGACATCAACAGTGGGTACATGACAGATGGAGGACTTAACCTATATATCTATGAAAGTCTGAA
1681	CCGAATACCAACAGACAGCAACTTCCCGGGACATCATCCAGAGAGGGTTTACGATGTGACAGTGGATGACAGACAGCTGGGATGACAGCACTTTCAGTGAAGCAGTGGTCTTCAGTACACCCCT
1801	TGATAACATCAGCATGACCTGAACACCCACATCCTCTGTCTAGCTTCTTACTCCAAACATCACCGTCCCTCTCAGGAAGAAATACTCAGCTGAGGACAGATTCAGAGAAACCGTCCACAC
1921	AGACGACACCTGGGATAGTCTCAGGAACTGAANAACCAAGAAAGATTTTACAGCCCATGCGGTGCGCAAGTGAAGACCTGTCTCTGAGACTTCTCTGAGACTTCTCTGAGAGACCCCGAGAA

Fig. 1

2041 GGACGGGACAGAAAGCTTCCCTGTCTGTTCACAGACAGGTTTCCCTGGAGAGAGGCATGTCTGCCCAAGAGGGGCCCATCTAGGCAGAAAGCTGGAAACAAGTGCACATCAAAACACCCCGG  
A G Q K A S L S V S Q T G S W R R G M S A Q G G A P S R Q K A G T S A L K T P G

2161 GAAACCCGATGATGCCAAGCTTCTGAGAAAGGAAAGCTCCCTTAAGAGATCACTCTACAAAGATCTCTTCAGATGCGAGGAAAGAGAGTGGAGATGAAGGAAAGAGCCGCCCTC  
K T D D A K A S E K G K A P L K G S S L Q R S P S D A G K S S G D E G K K P P S

2281 AGGCATTGGAAGATCGACTGCCACAGCTTCTTGGCTTTAAGAAACCAAGTGGAGTAGGCTCATCTGCCATGATCAACAGAGTGCAGCAACCATACAAAGTGGCTCTGCAACACATGGG  
G I G R S T A T S S F G F K K P S G V G S S A M I T S S G A T I T S G S A T L G

2401 TAAATTCCAAATCTGCTGCCATTGGCGGGAAGTCAATTCAGGGAGAAACACAGTTGGACGGTTCACAGATTCAGAGTATGATTTGCTGCTGATGTTAGCTCAAGAGACTACCCCTACA  
K I P K S A A I G G K S N A G R K T S L D G S Q N Q D V V L H V S S K T T L Q

2521 ATATCGCAGCTTGCCCGCCCTTCAAAATCCAGCACCAAGTGGCATTCCTGGCCGAGGAGGCCACAGATCCAGTACCAAGTATGATTCACAGTCCAGCAGCAAGTCTGCTGGGGCCAC  
Y R S L P R P S K S T S G I P G R G G H R S S T S I D S N V S S K S A G A T

2641 CACCTCGAATCTGAGAGAACCAACTAAATTTGGTCAAGGCGCTCGAGTCTCTCACCGTCACCAACAGACAGAAAGGAAAGTAGCAGTCTCAGATTCAGAAAGTGTTCCTTT  
T S K L R E P T K I G S G R S S P V T V N Q T T D K E K E K V A V S D S E S V S L

2761 GTCAGTTCCCCCAAATCCAGGCCCCACCTCTGCCAGCGCTGTGGTGCACAGGCTTCAGGCAGCCAGATCCAAAGTATCCAGATATTCGCTTCCCTCACCACATTTTCGAGGTTGTTGGTGC  
S G S P K S S P T S A S A C G A Q G L R Q P G S K Y P D I A S P T T F R R L F G A

2881 CAAGGAGGTGGCAATCTGCCCTCGACCTAATCTGAGGGTGTGAATCTTCTCAGTAATGCCAGCCCTAGTACCACATTAGCGCGGCAAGGAGTGTGGAGTCAACGTCGTCGCGG  
K A G G K S A S A P N T E G V K S S S V M P S P S T T L A R Q G S L E S P S G

3001 TACGGCAGCATGGGCAGTCTGGTGGCTTAAGCGCAGCAGCAGCCCTCTCTCAATAACCCCTCAGACTTAACCTACAGATGTTATAAGCTTTAAGTCACTCGTTGGCTCCAGGCCACG  
T G S M G S A G G L S G S S P L F N K P S D L T T D V I S L S H S L A S S P A

3121 ATCGGTTCACTCTTTCACATCAGGTGGTCTCGTGGGCTGCCAATATAGCAGTCTCTCTGACGACAGAGTACTCCAGCTTACCGAGTCCAGTCCACTAGCTCCACACGAGCTCTGA  
S V H S F T S G G L V W A A N M S S S A G S K D T P S Y Q S M T S L H T S S E

3241 GTCATTGACCTCCCTCAGCCATCATGGCTCTCTGAGCTTACACAGGCACTCAGAGGTHCCAGAGCCCTGCTTCATGAGAACGGGTAGTGTGAGATCTACTCTCTCAGAAAGCAT  
S I D L P L S H G S L S G L T T G T H E V Q S L L M R T G S V R S T L S E S M

3361 GCAGCTTGACAGAAATACACTTACCCAAAGGNCATAAGNATATACCCCATCTCTCGCAGGCCCAACCAAGMAGGGCAAGAGTGGTTGCGTTCATCTCTACTGGAGGGCTTCAGGA  
Q L D R N T L P K G L R Y T P S S R Q A N Q E E G K E W L R S H S T G G L Q D

3481 CACTGGCAACCACTCAGCTCTGGTTCCCTTCTGCCAIGTCTATCTCTTCAGCTGGAAATACCACTTCTTAACCTGGTGAGGCCCAACAAATTTGTCTCAATTTAACCTTCCCGGGCC  
T G N Q S P L V S P S A M S S S A A G K Y H F S N L V S P T N L S Q F N L P G P

3601 CAGCATGATCGGCTCAAAACAGCATCCCGCCAGACTCTTCTTCGATCTCTATGATGACTCCAGCTTTGTGGAGTGCCACTTCTCTGGAGGAAGACCTCGTGCCATCAGTCAATTC  
S M M R S N S I P A Q D S S F D L Y D D S Q L C G S A T S L E E R P R A I S H S

3721 GGGCTCATTCAGACAGCATGGAAGAAGTTCAATGGCTCTTCAATTACCTGGTTCAGCACTTCTTCTTACTCTACAGCTGAAGAAAGGCTCATTCAGAGCAAAATCCATATAACT  
G S F R D S M E E V H G S S L S L V S S T S S L Y S T A E E K A H S E Q I H K L

3841 GCGGACAGAGCTGGTTGCATCAAGAAAGTTGCTACCTCAGCTTTCAGCAATGCTCACCCTTGATAGCAGCTTTTGAAGAAGCTTAGGGAAATATGACTGGCCGATTCGA  
R R E L V A S Q E K V A T L T S Q L S A N A H L V A A F E K S L G N M T G R L Q

3961 AAGCTTAACATGACAGCGGAACAAAGGAATCTGAACCTTATAGAACTTAAGAGAAACCAATGAATGCTGAAGGCTCAGAAATCTGCTGCCCAAGCGGCTATTTCAGGGAGCACATGAATGG  
S L T M T A E Q K E S E L I E L R E T I E M L K A Q N S A A Q A A I Q G A L N G

Fig. 1 (cont'd)

44081	TCCAGACCATCCCTCCCAAAGATCTTCCGATCATAGAGACAGCATTCCTCTCTGAGTATCCAGCATATCCAGATGCCACAAGCCATTCACAGTGCACCAAGTGTAAATGANTGCCGACTCCAA P D H P P K D L R I R R Q H S S E S V S S I N S A T S H S S I G S G N D A D S K
44201	GAAGAAGAAAAGAAAAGAACTGGCTGAGAAAGTCTTCAAACAAGCCCTTGGGAGAGAAAGTCCACCAAGCCCTCTTCATACATCTCGACATTAAGAGAGCTTACTGATTCATCCCTTC K K K K K N W L R S S F K Q A F G K K S T K P P S H S D I E E L T D S S L P
44321	GGCATCCCCCAAGTTACCCCATANVGTCTGTGACATGAGCTCAGCATCCCAAGAGCCCTCACAATCTGCTTCACGGATCTGTGAATGCACAGAGAGCTGAGGCGAGAGATAATTTCTGCAGCT A S P K L P H N A G D C G S A S M K P S Q S A S A I C E C T E A E A E I I L Q L
44441	GAAAGCGAGCTCAGAGAAAAGGAATTAATAATTAAACGATATTTCCGGCTGGAGGCCCTCAGCTCTGCTCATCATCTTGTATCAGATCCGGAGAGCCATGAACCCGGATGCGAAGATGAATAATTGA K S E L R E K E L K L T D I R L E A L S S A H H L D Q I R E A M N R M Q N E I E
44561	AATATCTCAAAGCTGAAATATGACGGTGTGAAGCGAAGAACTGGTAAACAGCTAAGCCTACTCGGCCACCGTCAGAGATCTCTCAAGCAGCACCTCTCTTCATCTTCCAGGCGAGCTATTAGG I L K A E N D R L K A E T G N T A K P T R P P S E S S S T S S S S R Q S L G
44681	ACTTTCTTAACCAATTTGAACATCAGAGGGCTTTAGCTCAGATATTTTGTCTAGATGATGCTGTGTATGCAACTGGACATTAAGAGATGCGCGAGTGTGAAAATTAATGATCTCCATAAG L S L N N L N I T E A V S S D I L L D D A G D A T G H K D G R S V K I I V S I S
44801	CAAGGCTATGGTCGAGCAAGAACCTCAGGCATATTTGTATAGATCCATTTGGTGTGTAGTGAAGAAAACCAAGTGGGATGTCTTAGATGTGTAAATTAAGACGCTCTCTTTTAAGGA K G Y G R A K D Q K S Q A Y L I G S I G V S G K T K W D V L D G V I R R L F K E
44921	ATATGATTCGGAATGTATACATCCACTAGGCTTGGTCTGAGCTCTGACTGCACTTGTCTAGCTACTGTATAGGAGACTTAAATTAGATCCCAATACCTAGAGATGCTGTAATTTGCTGCTTGC Y V F R I D T S T S L G L S S D C I A S Y C I G D L I R S H N L E V P E L L P C
45041	TGGATACCTTTGTTGGAGATAATAACATCATCAGCTGTGAACCTCAAGGGGTAGAGAAAATAAGTTTGGACAGTTTGTGTTTGTATACGCTGATCTCTAAACCAATTAACCAAGAGGTACTT G Y L V G D N N I I T V N L K G V E N S L D S F V F D T L I P K P I T Q R Y F
45161	TAACTGTGTGAGGACATCAGAAATATATCTCTCAGGACCGAGTGGTACTTGGCAACAAACAACTTGTGCAATATGTAAATAACCAAAATCTGGAGGAAAAGAAAACAGA N L L M E H R I I L S G P S G T G K T Y L A N K L A E Y V I T K S G R K K T E
45281	GGATGCAATTCGCCACTTTTAACTGTGGACCAACAGTCAAGTAAAGMATTTGCAACAATATCTTAGCTAAACCTGGCTGAACAGTGCAGTGTGATTAATTAATGAGTGGAGCTCCCAAGTTGTAAT D A I A T F N V D H K S S K E L Q Y L A N L A E Q C S A D N N G V E L P V I
45401	AATCTTGTAATACTTCATCTGGCTCTCTGAGTGATATCTTCAATGGTTTCTCAATGTGTAATACAAATAATGTCCATATATTTTGGACAATAATTTTGGAGCAATGAAATCAGGGAGTTTCTTCATC I L D N L H V G S L S D I F N G F L N C K Y N K C P Y I I G T M N Q G V S S S
45521	ACCAATCTAGAGCTGACATCAATTTCAAGTGGGTATATGTGCAATCATACAGAACCCAGTGAAGGCTTTTATAGGCAGATATCTTCGAGAGAAACCTCATAGAGATAGAAATTTGAAAG P N L E L H H N F R W V L C A N H T E P V K G F L G R Y L R R K L I E I E I E R R
45641	GAAATTCGCAATTAATGACCTAGTCAAAATTAATAGATTGGATTCGGAAGAGCGTGGCATCATCTCAACAGTTTITTTGGAAACACACAGTTCTCTGACGTTTACCATTGGTCCCGGACTATT N I R N N D L V K I I D W I P K T W H H L N S F L E T H S S S D V T I G P R L F T
45761	CCTTCCTTGGCCCATGGATGTAGAAGGTTCTAGAGTATGGTTTCATGGATCTCTGGAACATATCTTTTAGTACCTTAATTTCTTGGAGGCGAGTGAAGAGGGGTCTTCAGATGTATGGGAACG L P C P M D V E G S R V W F M D L W N Y S L V P Y I L E A V R E G L Q M Y G K R
45881	CACACCATGGGAAGATCTCTCAAAGTGGTGCTTGACACATATCCATTTGGAGCTCAGCAACTCTGCCTCAGGAGAGGCCAGCCCTTACTTCAGCTGCGACCAAGAGATGTTGGGTATGCAAG T P W E D P S K W V L D T Y P W S S A T L P Q E S P A L L Q L R P E D V G Y E S

Fig. 1 (cont'd)

[illegible]

[illegible]

Fig. 1 (cont'd)

090324Z JUL 68 : 0000000000

5/124

## Human genomic sequence

```

1  GATCAGACTT TGAAGAGTGT TTGTACCATG CTAAAGTTTA CAGAATTTAT
51 TCCTGCTCTT TGAGGGTGCA TTGCAAATCC AGGCTAGAGG GAGAGATACC
101 AGTTAGGAXA GTACAGCAAT ACTCTACTGG GAAATGGTGA GGTGTTTCGT
151 GAAGACAATG GCAACACAGA TGAAGACATG CAGATGGAGG AAATAAAGAT
201 CCAGTTGAGC TTGTTGGCCA GTTGGATAGA GGTGAGGTT ATGCATGATG
251 GAGCAATCTA GGTTTTTGTC TTGGGTAGGT GTTTCATGA TAGTACTCAG
301 AATGAATCAT ATAGTTGTAC AGGTTGAATC CCACCCATGT TTGCACAATA
351 GAGTGACTGT CTAGCTGAAA TCCAGATGAC ACTCTGTATG CTAAGCTATG
401 CTTTCATGGAA CTGTATAAAG GCACTTGCTA CATAGGCTAG TGGCAGATCT
451 GGAAGTAACC TATATGGTAT ATAGGAAATG AGGTGGCTTT TGTATAAATC
501 CTACAGATAA ATTTTCATTC CTGATCCTAT TATTTTGACT CATGTTAGCC
551 CAAGAAGAGT ATTCAGTACT TCATATCCCT GAAGGTAAGA CAGAGTAGTA
601 TTAGATTCAC TATTTGGCAA ATAAAAGGGA TCAAGTCCTA AGATCAAGCT
651 GATGAATCAA CACCTCATAG GATATGTCCC AACCAATTAT ATGGCTTCCC
701 CTATAAATAA AATCTAGTTC TCTTCTCTGG AGAGGAACAG TGAAGAATAT
751 CATAACCTAT GCTACAAACT GCTTGAGTAG GAGCTACTTC TCTCCAAGGC
801 TTTATATCAT TCATTCTGGC AGGCCCTCT GTTTGTTCTC ACCAGCTCCT
851 GGGAAATTTA TTTCTCCTCT AGTGATATAA AAGCTCTCTG TTTGAGATGA
901 AGGGCTGCCC AGTTTATCAG ATCTGTATTA GTCTGTTCTC AGGCTGCTAA
951 TAAAGACATA CCTGAGACTG AGTAATTTAT GAAGGAAAGA GGTTTAATTG
1001 ACTCACAGTT CCACATGGCT GGGGAGGCCT CACAATCATG GCGAAAGACT
1051 AATAAGGAGC AAAGTCACAT CTTACATGGC TGCAGACAAG AGAGCATGTG
1101 CAGGGGAAC TCTCTCCATA AAACCATCAG ATCTTGTGAG ACTTGTTTAC
1151 TATTACAAGA ACAACAGACA GGAAAACCCG CCCCCTCAAT TCAATTACCT
1201 GCCACTGGGA CCCTCCCACA ACACATGGGG ATTATGAGAG CTACAATTCA
1251 AGATGAGATT TGGGTGGGGA TACCGCCAAA CCATATGAAG TTCTTTCTTT
1301 GTTACTGGGT ACCATATCCA TTCTGTTGAG GTTCTGAGCC TTTCCAGTTA
1351 CTGTAAC TCTATCTCCT GTCTGTGCTA AGACTCAGTG ACCTCTCTCT
1401 GCCTTGCTTC TGCTTTGTCC TGACCTTTC TGTGCATGCA CTCACTCTAG
1451 TTTGCCACC TGAGGTGAGA GATGGTCCAG ATTAGCAACA ACAATCTGTG
1501 GACTAAAATC CTCTTTAGGG AGGAAGCAAA ATTCAGATGG ATGTTACTAA
1551 ACAAAGCTCA GAAACAGAGA CCAGGGTGTG GGAAGTAAGG TAGTAGCCTG
1601 AGAGCAGCTG GCAGTGT TTT AGACCTGGAG GGAGGTTAGG TCATCAGCAA
1651 TGAGGAGACT GCCTGGAAAA TCCTAGAAAA TTAAGACATC TGGTCAGGCA
1701 AGGTCATATC ACCAGCACAC TTCCCTTTTC AAGTTGAATC CCTTTCCTCT

```

Fig. 2

6/ 124

1751 GTTAAGAGGA TTCAAGTGTC TTTCTTG CAT TTTGTCTTCT CTCTATATC  
 1801 CATGCTTGCA ATATAAGGAG ACAGCAGTTG GCTGTTTG TG CTAGAAAATA  
 1851 TAAATGGCCA TTTTGAAAGC ATGCCAGACA GGATCTGCGG CAAGTTTTCA  
 1901 ATGTTACTGC TGCCATCTGT TGTTCCTCAG TGCTGGGATG TGAATCTCTT  
 1951 GGCAAACATC TCTCTAATTC TGAACATCTT TTCACCCCCA TCTAGAGATA  
 2001 TTCACCTACT GAAGTGCCTT TTTAAAGCAA TGTTCCCTCAC CAAGGCGATG  
 2051 TTCTGAATGT TTTAAAATGG AAGAATCTGG AATGTTTTTA TTATAATACA  
 2101 TTTTGTATAT CCCAAAGCAA AAATCAATTT CTTCATGGTT AATACTTTTG  
 2151 TAATTTTGTT TTTAATAATA TTTTCCTTTT AAATATAAGA AATATTTTAT  
 2201 TGAATTAATA CTTTAATGTA GCTGTTTCAA GTAAGATAAA ACAGAACAGA  
 2251 TTACTGTTTT CAACCTTGTT CACAGTTAGC TCTGTAAC TA AGTTGTTGAG  
 2301 CTTTATCTAA GCTTTTTTAT TTTTACATAA CGTTTCCCTT TTCACTTAAC  
 2351 CTTGAAATTA TAGTAATTTG GGAACCTCTA TTCCTCTGAA AGAGAAAGCT  
 2401 AATGCCAAAG ATATTTCAAG GGAGAAAGAA GGTTTTTTAA AGGAGAGACA  
 2451 ATTCAGCTCA GACTTAATAG CTGTGATTGC TATTTATTAA GCAGAACGCC  
 2501 TATAACTAAA TTCTCAGATA TCCAAAAAAC AGCCTGTACA TTCTCAAAAG  
 2551 TGAAGATTAC ACATTTTCTA AGTTAAGGTA AAAGTTTTGT CTCTGTAGCA  
 2601 TCTTACTGAT TTCTATCTTC TCATTCTGCC TTAATAATGT CACTAAATAA  
 2651 ATGTTTGATG CACTAATACA TGAATAAAAC TATTCATGGT AATGATTCTT  
 2701 TAGAAACACA GCTAAGTTTT GTAATTTTGT TTTTAAAAA TTAAAAATTT  
 2751 AAATATAAAA ATGTTTTTAA AAGGCTTGAA TTTCTTGTTA AATGTACACA  
 2801 TTTTAAGTTG TAGGCTGTCT TTAATAATAA TCTCTCCACA CACTGTAGTA  
 2851 TTTAAACAT CATGATATTA CTATAAAACA TCAACAAATA GGGCAGTGGA  
 2901 AAACATGGTA ATCACTAAAA ATGCTCACAT GTCATATATT AAGACTTGAT  
 2951 AAGTAAACCA CAATAATAAA TAGAAAAGAA ATAGTTGTCT AAAAAGGGAT  
 3001 TCTCACCTTT CAAACCTTAC CATAAAAATG GAATATAAAA GAAGGAAGAG  
 3051 GAGGAGAAAT CAAATTATAT CATAAAAATTT TCTGGGCAA AATATTACAG  
 3101 AAGAAAATAA GAAAGATTTA TGGAGTTGAC TGAAACATTT TTGAATCCTA  
 3151 TACATAAAAA TATCGTTAAT TAAAAGGAAA AACAAAGAAA CAGATTTGGG  
 3201 AAATATTTGA AACTGGTTTT TTTT TAGCAT TTAATAATGT AATACAAATG  
 3251 GATTATTTAA ACTCCATTGC AAAAATACAC AAAGGACATT GACAATGTCT  
 3301 GGAAATAAAA TTAGCTAAGT AAGTTATAGA AAAACTCAGT CTCACAATTT  
 3351 GACAAATGTA ACTGAAAAC ATTAATATAA TTAGTAAC TA TTTTACATG  
 3401 TCAAAATTTT TGAATTACTA AAGGAAACCA CAATGCCTGA AAGTATCCAG  
 3451 GGTTTTTTTT TTTTTTTATA ATATTGGCAC TGTCATATGG GTGGCAGGAA

Fig. 2 (cont'd 1)



7/ 124

3501 TTGAAGTGAT GTTGTTCCTT CAGTTATTAA GTTGCATCTG CAGTGTTCCTCA  
 3551 AATGTCCAAA ACCTGTGAGT CAGTAATTCT CTTTTTGTAT ATTTATCCTA  
 3601 ATACAATAAT TCTAAACATA ATCTCAATAT ATATGTACAA AGTTATTTCAC  
 3651 TGCAGTGTTA CTTACAATAG TTAGAAAATT GTAAAATGCT TTATGCATCT  
 3701 TAAAATATAA ATTGTTGAAT ATATAATAGT CCATATGATA TAATTATATC  
 3751 ATTATTATAA ATAATGAATT AGAAAATAAT TTAAGAGCAT TAAAAATAATT  
 3801 ATAAGGTAAT ATGAAGTGAA TGAATAATGT ACAGATACTA TAATCAGCAG  
 3851 AGTGTTAACT AGGTAAATTT TTATGTGTGT ATATACTACT TCCTAAAAAT  
 3901 GACTTGACAG AAATCATCAA AATGCTAATG GTGGTTACTT CTGGGTGGGA  
 3951 ATACAGATGA TTTACTTTGT TCCTTTTATG TATTTCTGCA CTGCCCAGTC  
 4001 TTCCACAGTG AGCATATATT GGTTTTAA TTTATATAAG ATGGAAAAAG  
 4051 ATACCAAATG GTCTTCAATG AATCCTGGAG TTAACCTTCA TGTGTGTCAT  
 4101 ATGTTATATT CTAAACTTAT CACAAATAGA AGACTTTAA TCAACTTGTA  
 4151 CCTATTTCOA CTATATAACA GCATCTTTAA AATGAGCATT GAATTAAACT  
 4201 ACCAAAACCA ACCATCATGA GGATTATTCA AGTAATGTGT TTAAACAAAA  
 4251 GAATTTGTAA TAAATTTACT TTATCTCCTT TGTGATTTC ACCCCATTAA  
 4301 AAAAAATAGA TGTTTCTACT CTCCTTCAGA TATCATTAAG ACATAAACTT  
 4351 GTGCCTGACT GCATAAATCC CTTTAAACT AATATCACTT ATTACGTTTA  
 4401 ACTAAGTCTA CCTAGGGCTT CCTTGTATAA AGAACAAGAG CTTTCCATTT  
 4451 TTTGTTTACC TAGCCCTTTC TGATGCCACG ACAGAATAGC TGTAATCTT  
 4501 CATTATTTAT ATTCTAGAGA AAATAAAAGC AAATAAAAG GTCAGTGTAT  
 4551 AAAGTTTATT GGTGTTCTC TTTACTCAA ACCCACATGG TATTAATGTT  
 4601 AGTCTCTATG AATATTTCAT GGATAAAATC AGAGCATTAAG GTGCATACTA  
 4651 AAAACAATAA GAATGGAAAG ACTTTAACCT TATGTTTATA TGAATTTCTA  
 4701 GGTTATCAAG AAGTTTATAG GCTATAGGCT ATAAAGTCTT AGGCTATGAT  
 4751 ATAGTAACCT AATGTAGACT TCCCTTGATA CATGAAAATA ATGGTACTAA  
 4801 GTACAAACAG AAGATGAGCT TAAAATTATT CTTTGAGTCC TCTTGATGGA  
 4851 TTTTTTCCCC CACACTTCC CAAAATGT TTTATGCCTA TATTGTAGGA  
 4901 GACCATGCAA GAGACCTAGA GTCTCTTTTT CTTTCATCAC TTTCCAATCA  
 4951 ACAGCAAATC CTATCATTTT TACCACAAA TATATCTTGA AACTCCCTTC  
 5001 TTTTGATTTA CTTGTAACCT CCCATCAAAA ACTGAAGAGT GTCACAATAC  
 5051 TTCATTAAAGT TCCCTACTTG CACTCTACCT TTAATATATT TGTAGCACTA  
 5101 AAATGTTTTT AAAACATATA TCTGCTTATG TCATTTTACT GCTCAATACT  
 5151 ATCTGATTTT CTATTGCACT TCTAAGATAC TCTAATTTCT TAGCACTCTA  
 5201 TATAAAATCC TTTAAGGGCT TCCCTGCTCA CCTTTTCAGA CTCAGAACTA  
 5251 TGTATTTCTT TTTGCCTGCT GTACTTGATC CACTGGATTC TTGATTTTGG

8/124

5301 TTACTTCCAG GTTTTACAC TTATTTTAC AATAAATGTG AAATACCCTT  
 5351 TTTGACAATA TCTACAAATA TTTCTTATTT GTCTTTATTG CTCTTTCCTG  
 5401 TAATGTTTAG TCTTCATTTT CCTGATAATG GCTATCTAAA GTTATCTCCT  
 5451 CAAAGAAGCA GTTATTTATT CACCCAAATC TTCTAGTCCT TCTCTGGAGT  
 5501 TTTCTTCTCA CTTCATTCCC TTGGTTTTTG CCACAATTG TAATAATTTG  
 5551 CAATTTGGAG TGTTAGAATG AGGGAATAAA TCACAGGTAA TGACTATAGT  
 5601 TTGTGACTAT GTAAGATTGG ATTCGTTATT GATTTATTCC ACAAACACTG  
 5651 AGGCACTGCA TTTAGCCAAA TGCCAATCTT GGGCAGTGAG ACTCTGAAAG  
 5701 AGAATCTGCT TCCCCACCA TAAACTACAA AGTGAAACAA CTCAGAATGT  
 5751 ACATAAATTA CAGAATGAAA GCACACTAGA AGTAAACACA GATGTGGAAG  
 5801 AGGTAAAGTG TCCTTGAAAA TCATGGAAAG ATTCATAAAG GGAATGACAT  
 5851 TTCAACTGGA TTCTAAACCA GTTATTCAAG CTCCACAAGG TTGCACAGTA  
 5901 AATGAGCAGT GGCAGGATGA CATACTTAG AAAGTAAAG GAATCTTTTT  
 5951 TAAACTGCTA TAAAAATCAT TACATATACA TTTTGTAGGT CGAGAGTAAG  
 6001 GTATTTAACA TAAATCATT TTAGTATATC AGTGTTTATA TAGACTTAGG  
 6051 TTTTCTCAT TTAACCTC TTTTAATGAC TTGTGCTTTT CTTCATGGTA  
 6101 ATAAACATT TTCCAGGAA GTGCTGAATA AATCTTCTT GAAATACGTT  
 6151 TTATTGCTTT CTATCAATGA CCCTGAAGTA ATACAGAATT TACTTTCAG  
 6201 CGGTTGCAAT GCTCAAACCT GACAGGTAAT GCACTGTGTT TGCTGATATA  
 6251 AGAGGTATGA TGTAGGGCTA AGTGGTTTTG TGCTCATTTA GCTTTCAGGA  
 6301 GAAAATAATT GACTTAACAT TTTGATACTA AAACCCAAAG CCTAACAGTT  
 6351 AATTCTTGGT ATTTTAAATT ATTATTGCAA AGATTATTGT GCCGAATAAT  
 6401 ATGAAAATAT TTTATATAAT ATTTAAAAAG TATATCTCTT TCTTGGTATT  
 6451 ATTTAAATTA CCATAAAAAAT GTGCGAAAAA GTTATACTGA AATGTGATAG  
 6501 GATCTTTTAA AAGTGGTGCC TTGATTTTGT TAAGTGTTAC CTAGTTTTC  
 6551 TCTGAAACA AGAAACATAC CCAGAAGTTT TCACGAAATG GTCTCATGAA  
 6601 TATCTAAGGT TAGTCCGTAG TCTCATCTGA GACAAGGAAA GTCCCTTCCA  
 6651 CTATGAGCCT GTAAATCAC AAGCAAGCTA GTTACTTCCT AGATACAATG  
 6701 GGAGTACTGG TATTGGGTAA ACACAGCTGT TTCAAATGGG AGAAATTGGC  
 6751 CAAAATTAAT GGGTTACAGG GCATGCAATT CCGAAATCCA TCTGGGCAGT  
 6801 CAAATTGTAA AACTCCAAAA TGATXTCTTT TGACTCCATG TXTCACATCC  
 6851 AGGACATGCT GAXGCAAGAG ATAGGTTCCC ATAATCTTTG GCAGCTCTGC  
 6901 CCCTGTGGCT TTGCAGGGTA TATCACCCCT CCCAGCTGCT TTCACAGGCT  
 6951 GGCATTGAGT GTCTGTGGCT TTCCCAGGAA CAAGGTGCAA GCTGTTGGTG  
 7001 GATCTACCAT TCTGGGTTT GGAGGATGAT GGCCCTCTTC TCATAGCTCC

Fig. 2 (cont'd 3)

9/124

7051 ACTAGGCCGT GCTCCAGTAG AGACTCTGTG GGGGCTCTGA CCCCAGATTT  
 7101 CCCTCCTGCA CTGCCCTAGC AGAGATTCTT CATGAGGGCC GTGCCCCTGC  
 7151 AGAAACTCTT TTCCTGGGCA TCCAGGCATT TCCATACATC TGAAATCTAG  
 7201 GTGGAGGTTT CCAAACCTCG ATTCCTAATT TCTGTGCACC TGCAGGCTCT  
 7251 CTACCACGTG GAAGCTGCCA AGGTTTGGGG CTTGCACCC TGTAAACCAC  
 7301 AGGCTGAGCT ATACCTTGGC CCCTTTTAGC AATGGCTGGA GTGACTGGGA  
 7351 CACAGGGCAC CAAGTCTCTA GGCTGCACAC AGTATGGGCA CCCTGGGCCC  
 7401 AGCCCTCAAA ATCATTTTTT CCTCCTAGGC TTCTGGATCA GTGAAGGGTG  
 7451 GGGCTGCCAT GAAGACCTAT GACATGCCCT GGAGACATTT TCCCCATGTG  
 7501 CTTGGGGATT AACACTGGCT CTTGTCTACT TATGCAGATT TCTGCAGCCA  
 7551 GCTGAATTTT TCCTCAAAAA ATGGGTTTTT CTTTCTACT GCATTGTCAG  
 7601 GCTGCAAATT TTCTGAACTT TTATGCTGTT TCCCTTTTAA AATGCGATGC  
 7651 TCTAACAACA CCCGTCACCT CTTGAATGCT TTGCTGCTTA GAAATTTCTT  
 7701 CTGTCAGATA CCCTAAATCA TCTCTCTCAA GTTCAGAGTT CCACAAATCT  
 7751 CTAGGGCAGG GGCAAAATGC CACCAGTCTC TTTGCTAAAA CATAACAAGA  
 7801 GTCGCCTTTG CTCCAGTTCT CAGCAAGTTC CTCATCTCCA TCCGAGACAA  
 7851 CCTCAGCCTG GTCCTTATTG TTTATATCAC TATAAAAATT TTTGTCAAAG  
 7901 CCATTCAACA AGTCTCTACT CCAAACTTTC CCACATTTTC CTGTCTTCTT  
 7951 CTGAGCCCTC CAAATTGTTC CAGCCTCTGC CTGATACACA GTCCCAAAGT  
 8001 TACTTCCACA TTTTGGGATA TCTTTTCAGC AATGCCCGC TCTACTGGTA  
 8051 CCAACTTACT TTGTAGTCC GTTTTCACAC TGTGATAAA GACATACCCA  
 8101 AGACTGGAAG GAAAAAAGG TTTAATTGGA CTTACAGTTC CACATGGCTA  
 8151 GGGAGGCTTC ACAATCATGG CAGGAGGCAA AAGGCATTTT TTACATGATG  
 8201 GCAGCAAGAG AAAATGAGGA AGATGCAAAC GCAGAAATCC CTGATAAAAC  
 8251 CATCGGACCT TGTAAGACTT ATTCACTACC ACTAGGACAG TATGGGTGAT  
 8301 ACCACCCCCA TGATTCAAAT GATCTCCAAC CAGGTGCCTC CCACAACACA  
 8351 TGGGAATTAT GGGAATACAA TTCAAGATGA GATTTGGGTA GGGACACAGA  
 8401 GCCAAACTAT ATCACATGGA TTTCTTATAC TTTTGCTTTT AATAACACAA  
 8451 ACAAAAAAAT ACATCATTA AAGGTTAGAA GTGAGAAGGT GTTTTATGG  
 8501 AAATCAAAAA TAATATCACC TTAGTGAACA GTATTCTTAT GATTGTAGTT  
 8551 GAATTAGAGA GCAGAATACA TCTAGAAGAT TCAGTAGTAA GCATGTTTCT  
 8601 TCGATTAATG GAAAATTTGA ATAGCCTAGC TGATTGAGAT TGAGGTACT  
 8651 ATTAAATGCC TGAAGTATAA GAGTTGGTTG TTTATGTAAA CAAAATATCT  
 8701 GTTTTACATG TACATGTGTA AGTAGGACTG TTGAGCCCCA GTAACATGAA  
 8751 ATATCAAAGA GCATGACTCG AATACCTGCC ATATGAAGTG CTATTACATC  
 8801 AAAAAAGAGG CGTGTGCTGA AAAATTACCT ACAAATGGCA TTTTCTCAA

Fig. 2 (cont'd 4)

10/ 124

8851 ATCAATTTTA AATCTTCAGA ATTTTCATTTT AATAATTGTT TAGTTAATAT  
 8901 TTCAGAAATCC CTCATCATAA AAAGCAGGCA AAAGGCAAAA GTCCTTGAAT  
 8951 GTATAACACA TTTGTTTTCA AACAAGCCTG CCTCTAACTG TGAATCCAGG  
 9001 AGTGAATCCA GAACTACAAA TTAACATAAGA TTGGCCCCAT CGAGTTACTG  
 9051 AACGTTAAAA ATCTAAAAAC TAAAAGGCAT GCCTCAACAA TTATTTTCTT  
 9101 CTTGGAATCA TTAATTAACC TATGTGTATC CAAACAATAA TCTTCCAGCA  
 9151 GTTTCGCTAG CTACATTTTT AATTACTTAA TATCATGTAA AATTGTTTTT  
 9201 ATTATTGTTC AGTTCTGAAT TTTGACATAT GCATCAAGCC ATGCAACTGC  
 9251 TACCACAGTC TTCCTGATCA CTGATCTGTT CTAAATCTCT ATAGCATTTT  
 9301 TCCTTTTCTT AAATGTTGCA TAAATAAAAC CATACCTTAT GTGGCCTTTT  
 9351 GAATCTGGCA TCTTTAACTT AATGCGCTG AAATTAATCT ATGTCATTTT  
 9401 ATGTATCAAT GGCTCAATCT TTTTAATTGT TAAGAAAAAA TGTATGCTGG  
 9451 GATAAATATC TTTCTAAATG AGTTTTTGTT CACAATGCTG AGTGTGTTGTT  
 9501 TAGGATAGAG TCCTAGAAAT GGTATCACTA GGTCAAACAT TCAAATAATT  
 9551 TTAAATATT TGATACATAT TGCCAAATAA TCTCAAATTT TTTACCAATA  
 9601 TACATTTATG ACAGTATGGG ATAAATGTGT CTTTCTTATA CCAACTGACA  
 9651 ACATTAATGA TAATACATAA AATATTCTTT GCTAATTTGA TGGGACAGAA  
 9701 ATGTTATATC CTTATTAGCA TTTTATTATT GTGGTTGAAT GACTGTACTG  
 9751 TACAGCCAGA GATATTTGGT TCAAAATCCA TCTTCATTAT TTAGTGTATG  
 9801 TGAAAATTTA GGTGAGCTAT TTAATCTCTT GATGCCTTAG TCTCCTAATC  
 9851 TATAAAGTGG GGATAATTGT ACCAATCATA TTAGGTTCCCT GTGAGAATTA  
 9901 ACTGAATTAC TATAGAAAAT GCTTAGAATG GTATCTAGTC ACCAGGAAGG  
 9951 ACTCTCTCTG TATTACTTGT TTATTATCTA ACACGTTTAA TTATTAATGA  
 10001 AGCTCAGTTT CGTTATATGC TTGGGATATT TGAAACTTTT CTTAGTGAAT  
 10051 TTTCCAATAA AATTATTTGT CTATTTTCTT ATGGACAAGT TGGTATTATT  
 10101 CTTACTGGTT TGTTTCAGGT TCAGTTAGTA AGAATTTTAA GGATTTTCTA  
 10151 TCACATTTTA GCAAACTTTT TCTGCATTTT ATCTTTTTTC TTTCAGATAA  
 10201 TGTTTGCAAA ATGTAAAAAA AACAAAAGGT TTCTTCATCA AGTTGGTATC  
 10251 TTTATCTTTT TTATTGCTTT GTGATTTGAA AATTCTTGTC CTGAGAACCA  
 10301 AAATATATAT TTGATGAAAT AGTTCTCTTC TTTTACTCAT TCTGAAGTCA  
 10351 TTGGAATTGA ATTTGGCATA TGATATAAAT CCTAATTTTA TATTTTATGA  
 10401 TATTCAAAAT TTCTAACAAA TATTTACTTA ATAATCTAAT CCAGGTTTCT  
 10451 ATTGTTTCTT CTGTTTCCTT TATAATGCTT TTTCTGAAGT TATTTTTCCT  
 10501 AGACTTAAAT ATTAGTATAA TATTATCATA GAGGAAAAAA TATCTGTTAG  
 10551 CTATGAATAA AAGGCTTTCA TCTTATTGTT GCATTAATAT ATTTAAATGT

Fig. 2 (cont'd 5)

11/ 124

10601 AGAGAGCATA CAGATTAGCA AAGAAAAAGT ATAATTGCCT TTTTATATAG  
 10651 TTGACATGAA CATGTATAAA GAAAAACCA AAAAATCAAT AAAACAACATA  
 10701 GAACTTATTA GTGAATTTAG CAAGATCATA GCATACAAAG CCAAGATTCA  
 10751 AAATTCCATT TTATTTATCT ACTAACAAAA AATATTTGAA ATTTGAAAAAT  
 10801 TTAAATATGC CATTTACAAT AACATCAAAA TATTGAACAA TAAAGTATTT  
 10851 AGGAATTTAT AAAATGAAAT CTCCTATACC AGGAATTACA GACCATTGCT  
 10901 GAAATAAATG AAAGAAGACC AATATATGTG AAGAGATACT CATTTGTGGA  
 10951 TTGAGAGACA ATATTGTTAA AGTATCAGTA TTTCCCAAAT TAATCAATAG  
 11001 ATTCAATATA ATGGTGAACA GAACACCAGA AGATGTTCTG TCGAAGCTGA  
 11051 CAAGCTATTT CTATAATTCA AATGGAAATG CAAAAGGCAG TCACTGCCAA  
 11101 CACCAGCATG GACTGTCTGG GTTCCAGTAG GTTACTTCAC TACTGCCTCT  
 11151 TCTGTCAGCC ACATCACGAC AGCTGCCCAG AAGCCAGAGA AACTCCTCAC  
 11201 ACCTGGCCCA CTGCTGCAGC TACCAGCATC CAGGCAAGCC ACCATCAGCC  
 11251 CACTGGTAAC TGCCAACAGA GGTACCACTG TACACTACCC TGGGGAACAA  
 11301 AGATAGGCAT GTAGTCAGCC CACCTCTGCC ACCACTAGGG CCTGAAGCCT  
 11351 GGCCACCTG ACACTGCAGT CCTCAGCACA GCTTCATCAC AGCTTCTGTT  
 11401 AATAACCACA CCCTAACCTA CCAAGGAAAT CACAAATGTC ACTGACACTG  
 11451 TTTGTAGCCA AAGAAATCAT AGAGAGACTA CATTACTGCA CACACCCATA  
 11501 ATCAAAGCCA CAGTACCCTA TCCAGACAAC ATCACAGGTA TATCTAAAGG  
 11551 AAAAAATTTT CCCATATGAA AGCGAATTCA AATATAGGAA GAAGCGACTG  
 11601 TTACAACAGA TATGCAGATA AAGCTTCAAC AATATCCTAC ATTCAACCAG  
 11651 AAGAAAGAAT CTCAGAAGGT AAAGACAGGT CTTCTGAAAT AATCTAGTCA  
 11701 GACAAAATTA AAAGAGAATA ATCAAATCCT TCCTGACATT TGGGATAACA  
 11751 TTAAAGTGAC CAAATATACG AATTATAGAT ACCCCTGAGA GTGAAAAGAC  
 11801 AAAGAAAAGA TTAGAAAACC CACTTAATTA AATAATATAT GAAAACCTCC  
 11851 TAAGTCTAGC AAGAGTTTGA GATATTTGGG ATGCAGGAGG CTCAATGGTC  
 11901 CCCAGGCCGA TAAAACGCAA AAAGGTCTTA TACACAGCAC ATTACAATCA  
 11951 GACTGTTTAA AGTCAAAGAT AAGGAATAAA TTCTAAAAAC AGCAAGAGAA  
 12001 AGTGTATGAT AACCTATGAA GTAAACCTTA TCAGACTGAC AGCAAATTTT  
 12051 TGGCAGAAAC TTTACAGGCC AGAAAGAATA GGACAATATA TTCAAAGTGC  
 12101 TTAAAGAAAA AAAAACTAT CAGCCTTAAA TACTATAGCC CACAAAATTA  
 12151 TCCTTCATAA ATGAAGGAGA AATAAAAGGT TTCCCAGACA CGAAAATGCT  
 12201 GAGGTAGTTT GTTACTACTA GACTGGACCT ACAATAAATG CTCAAGGGAG  
 12251 GTCTGGAAAC TGGTAGTGAA AGGACGACAT TTATCATCAT GAAAATACAT  
 12301 GAAAGTATAA AACTCCCTGG TAAGCAACTA AAGGGAGGTA TCAAATGTTA  
 12351 CCACCAGAGA AATCTAACTA ACCACAATGA CAAACAATAA GGGAAAAAGA

Fig. 2 (cont'd 6)

12/124

12401 AAGGAACAAA AATATATAAG ACAACAAATA AACAAACAATA TAACAGGAAG  
 12451 CCTCACATAT CAGTAATCAC TTTGAATGTA AATGAATTAC ATTCTCCACC  
 12501 TAAACGTTAT GAAATGCCTG AATGATAAAA CTATATGATC CAAATATATG  
 12551 CTGATTACAA GAAACTTACC AGGCAGACAT ACATAGGCTG AAAGTAAAAG  
 12601 AATGGTAAAA GATATTCCTT GCAAATGGAA AGCAATAGTG AGCAGGAGTA  
 12651 GCTATACTTA AATTAGATCA TACAGACTTT AAGTCAAAAA GAGTAAAATA  
 12701 AAAAAGACAA AGGATGTTAT TATATAATGA TGAGATTAAC CCAGCAATGG  
 12751 GAAATAACAA CTCTAAATGT ATATGCATTC AACACTAGAG AACTCAGATC  
 12801 CACAAAGCAA ATATTAGACC TAAAGAGAGA AATAGACTGC AATACAGTAA  
 12851 TAGTGGAGAA CTTCAACACT CCACTTTCAG TATTAGACAG ATAATCTAGG  
 12901 CAAAAAATCA ACCAGTAAAT TTTAGATTTA AACTAGATTT TAGACCAAAT  
 12951 GGACCTAACA GACATTTACA AAACATTCCTA TCCAACCACT GCAAAATGAA  
 13001 ATTTGTGTCA TCAGCACATG AAACAATGTC CAAGATAGAC CACCATATGT  
 13051 TAGGCCACAA ATCATGTCTC AGCAATTTTT TAAAAGTTGA AATCATATCA  
 13101 CATATCTTCT CAGACCACTG TTGAATAATG CTAGAAATCA ATGCCAAGAA  
 13151 TAACGTTGGA AACTATACAA ATACATGCAG ATTAAACAAC ATGTTCTTGG  
 13201 TTGATCACTG GGACAATAAG GAAATTAAGC TGAAAATCAA AAAATTCTTG  
 13251 TAACAAATAA AGATTGAAAC ATAACATATC AAAACCAGTG GCATACAGCA  
 13301 AAAGCAGTGC TAAGAGGGAA GTTTATAGCA ATAAATGCTT AACTGAAAA  
 13351 AGTAGAAATA TTTTAAAATT AGCAACCTAA CAATGTGCCT GAAGAACTA  
 13401 AAAAATCAAG AACAAATCAA ACCCAAAATC AGCAGAAGAA ACACAAAAAT  
 13451 AAAGATCAGA AAAGAATAA ATCAAATAGA GACTAAAAAA ATACAAATGA  
 13501 TTAACAAAAC TAAAATTTGG TTATTCAACA AGATAAATAA AATTGATAAA  
 13551 CCGCTAGATA GACTAAACAA GGAAAAAGAA TATCCAAATA AACACAATCA  
 13601 AAAACGATAA AGGAGACATT ACAACAGATG CCACAGAAAT AAAAAGGATC  
 13651 ATCAGAGACT ATTATTAACA ACTATATGCT GAAAAATGGA AAATATAGAG  
 13701 AAATAGATAA ATTCCTAGAA ACTTACAACC TACCAAGCTG TTGCATCAGG  
 13751 AAGAAATAGA AAACCTGAAC ATATCAGTAA TGATTAGCAA AATTGAATCA  
 13801 GTAATAAAAA ACATCTCCCA ACTCTTTTAA AGCTTTGGAC CAAATAGCAT  
 13851 CACAGCCTAA TTCTACCAAT CATGCAAAGA AGAATACCAG TCTTCTTGAT  
 13901 GCTATTACAA TAAATCAGAG GAAGGAATTC TCTCTGGCTC ATTCTACATG  
 13951 ACCAGTGTCA CCTTGAAACC AAAACCTGAC AAGGACACCA CAAAAAGAAA  
 14001 ACTACAGGCC AATAACCATG ATGAACACAG ATGCAAAAAT CATTAACAAA  
 14051 ATACTGGCAA ACGGAATCCA ACAGCACATC AAAAAAATAA TATACCACAA  
 14101 TCCAGAGGGT TTGTATCAAG GATACAAGTA TGACTCAATG TAAATAAATC

Fig. 2 (cont'd 7)

14151 AATAAACATG ATAAGCATCT TCACAGAATA TAAGACAAAT GAATATATGA  
14201 TCATCTCAAT AGATGCAGAA AAAAATTTTT GATAAATTTT AACATCTCTT  
14251 CATGAAAAAA ATCTCTAAAA CTCAGCATAG AAGAAACATA CCTCAATATA  
14301 ATAAAGGCCA TATGTGACAA ACTCAGAGCT AATATCATAC AGAATGGGGC  
14351 AAAGTTTAAA GACTTTCCTC TAAGAACTGG AACAAGACAA GGATGCAAAC  
14401 TCTCACCACCT CCTATCCACA TAGTACTAGA AGTCCTAGCC AAAACAATCA  
14451 GACAAGCAAA AGAAATAAAA AGTATCTAAA TTGAGAAGAG CAAGTAACAT  
14501 TGTTCTCTTT TGCTGATGAT ATGGTTTTGT ATCTGGAAAA TACTAAAAAC  
14551 TCCAGCAAAA ACCTCTTAGA TTTGATTAAT TAATTTAGTA AAGTTTCAGG  
14601 ATACAAAATA AAAATACAAA AGTCAGTAGC ATTTCTATGC CCCAATAATA  
14651 AAATAGCTAG GAAAGAAATC AAGAAAGTGA TCCCATTTAA ATTAGCTACA  
14701 AAAAATTAAA ATACCTGGGA ATAAATCAAG GAAGTTAAAG ATCTCTGCAC  
14751 AAAACTACAA AACACTGATG AAAGAAATTA AGGATTAAAC AAACAAATTG  
14801 AGAAACATCC CATGTTTATG GATCAAAAAGA ATTAATATCA TTAAATGAC  
14851 CATACTTCCC AAAGCAATTT CCACATTCAA TGCAATTTCT ACCAAATTAC  
14901 CAATGTCATA TTTTCATAGAA TTAGAATAAT CCTAAATTA GTATGGAATG  
14951 AGAACAGAGC CCAAATAGCC AAAGCAATTC TGAACATAAA GAACAAATCT  
15001 GGTCTGACT TAATCACTAT GCAATCTATG CATGTAACAA AATTGAACAT  
15051 GGATTTTATC AATTTGTACA AATAAAAAAA TGTAAAAAA GAACAAAGCT  
15101 GGAGGCTATA GTAGCCAAAA CAGCATGGTA TTTTGTAGACA AATGGAATGG  
15151 AATAGAAAGC TCAGAAATAA AGCCATATAT ATATATTGTG TGTGTGTGTG  
15201 TGTGTATACA CACATACATG TATATATAAT GTGTACATAT AATGTTTCT  
15251 ACATGTTCTA ATATTTATAT TCCATTCCAT TATACATATT CCATTTCTGT  
15301 ATATAGGTTA TATAGAATTG GAAGACTATC TGCCATTAAA AAGAATGAAA  
15351 TCCTGTGATT TGCAGCAACA TGGTTGAAAC TGGAGTTCAT TATCTTAAGT  
15401 GAAATAATCT AGGCACAAAA AGATAAATAT CACATGTTCT CACTTATATG  
15451 TGGGAGCTAA TAAC'TTGATT ACATGAAGGT GGAGAATGGA AAGGTAGGTA  
15501 GGAAACAGAG ACTGGAAAGG ATGAATGGAG GGTAGGAGG AAGGTGAAGA  
15551 GAAGAGAGTT AAAAGGTGTA AACATATAGT TAAAAGAAAT AAATTCAATG  
15601 CTTGATAGCA GAGTACAGTG ACTACAGTTA ACAAATGTA TTATACTCAG  
15651 GTGATGAACA CCTAAATACT TGATCACTAT GCAATTATAT ACGTGTAACA  
15701 AAATCACTAT GCAC'TATATA CGTGTA AAAAT TAAATGCGTA CAAATAAAAA  
15751 TAATAAAATA CTAATCCAGT ATCATTTACT GACAATGTTA ACTCAGGTGG  
15801 ATAGGCATTA AGTCAATACT ACTATAAGAA CCAC'TCTTG TTTATGTTAA  
15851 TGCCATATAG AATGAAATAA AATTCATAA AATCCAAAAA ATTAGAAAAA  
15901 CTATCAAAAC TCAATAATAT TAAGACAACC CAATAAAAAAT GTGGTCAAAG

14 / 124

15951	GATTTGAACA	TACATGTCAC	CAAAAAATAT	ATTCAAATTT	CCAATAAATA
16001	CATGTAACAA	TGTTTCGACAT	CGTTAGTCAT	CAGAGAAATA	CAAAATAAAA
16051	TGGTAATGAG	ATACTACTAG	ATAGGCTTTT	ACAGAGACTG	ACAATACCAA
16101	GTATTGACAA	GGATATGGAG	CAACTGAAAT	TCTCATTCCT	TGTGGTAAGA
16151	ATGTACAATT	ATATAACCAC	ATTGAAAAAA	CAAGTTTTCa	GTTTCTTTAT
16201	TCACCCAAAA	TATATGTCTT	TTGGAAAAAA	TTTTTTCAG	TCTGTGGGTT
16251	GTCTTCTCAT	TCTCTTGATA	TATGTCTTTT	CAAAGAGGCT	GAGCTTTACT
16301	TTAGACAGTG	GTCATCAAAG	TGTGTATATT	TGTGTTTTTA	TAATTTATAT
16351	GCATATATTc	CTGTGAAAAG	ATACTGTATG	CATTGTTCAA	CATGTACAAA
16401	TATAAGAAAG	ATATAGTAAA	GAAATATATA	TTTCTAAATT	TATAAATGTA
16451	TTTATTGGTG	TTCACCGTTG	CAAACTAAAT	AATCTACGTT	GGCTAATTTA
16501	AGGAATTAAA	CTATAGTAGA	AGGTTCTCAT	TTATTGGGAT	GATTAGAACC
16551	AGCCTTTTTG	CAGGCTATTA	GCGAATCATA	GCACTAGGGC	TTCACTGCTA
16601	CCTCCACTGA	CACCTCTGAC	ACTTGAAACT	TGAGGCCAGA	TATCTGCCCA
16651	TGCTGATAGA	AAACAACCTGA	ATAATTTAAT	TTGCTAGATA	ATAGAAAAGA
16701	ATCAAATGAC	TCTGCCACAT	TGCTTGCCAG	AAGATTGTTT	TTCTCATTTG
16751	TGACCTCTTG	CCTATAAATG	ATAGATAGTC	CCTGTGCTGC	ATGCTATAGG
16801	TGTTCGTAAG	AGAGTCTGGG	AATGTGAGCT	TTTTATATCC	TATTTTTGGG
16851	TGGTAAAGGT	CATTCTATTA	GTCTGTTCTT	AAACTGCTAA	TGAAGACATA
16901	CCCCAAATTG	GGTACTTTAT	GAAAGAAAGA	GGTTTAATTG	ACTCACAGTT
16951	CAACATGACT	GGGGAGGCCT	AAGGAAAGTT	ATAATCATGG	GGGAAGGGGA
17001	AGCACACATG	TCCTTCACAT	GGTAGCAGGA	AGGATAATGA	GTAAAGGGGG
17051	GAAAAGCCCC	TTATAAAACT	ATCAAATCCC	ATGAGAACTC	ACTCTCACAA
17101	GAACACAATT	AGAGTAACTG	CCCCCATGAC	TCAATTACTT	CCCACCAGGT
17151	CCCTCCCACA	ACACATGGGG	CTTATGGGAA	CTACAATTCA	AGATGAGATT
17201	TGGGTGGGGA	CACAGCCACA	CCATTTcATT	CCACCTCTGA	CCCCTCCCAA
17251	ATCTCGTGTT	CTCACAATTc	AAATACAATC	ATGCCCTTCC	AACAGTCCCC
17301	CCAAAGTCTT	AACACATTTc	AGTATTAACA	CAAAAGTCCA	AGTCCAAAGT
17351	CTAATCTGAG	ACAAGGCAAG	TCCCTTCTGC	CTATGAGCCT	GTAAATTCGA
17401	AAGCAAGTTA	GCTACTTCCT	AGATACAATA	GGGTCACAGT	CATTGGGTAA
17451	ATACACACAT	TCCAAACGGG	AGGAATTGAC	CAAAACCAAG	GGGCTACAGG
17501	CCTCATGGAG	GTCCAAAATC	CAATAGGGCC	ATTGTTAAAC	CTTAAAGTTT
17551	CAAAATTATC	TCCTTTGACT	TCATATCTCA	CGTCTAGGTC	ATGATTATGC
17601	AAGAGGTGGG	CTCCCAcAGC	TTTGGGCAGC	TCTGCCTCTG	TGGCTTTGCA
17651	GGGTACAGCC	CCACTCCAGG	CTGCTTTTAC	AAGCTAGTGT	TGAGTGCCTc

Fig. 2 (cont'd 9)



15/124

17701 CAGCTTTTCC AGGCACATGG GTGCAAGCTG TAGGTGGATC TACCATTCTG  
17751 TGGTCTGGAG GATGGTGGCC TTCATCTCAC AGATCCACTA GGCAGTACCC  
17801 CAGTGGGGAC TCTGTGTGGG GGCTCTGATC CCACATTTCC CTTCCACACT  
17851 GCCCTAGCAG AGGTTCACCA TGAGGGCTCC ACCCCTGCAG CAAACTTCTG  
17901 CCTGAACATC CAAGCATTTT CTTACATCCT CTGGAATCTA GGCGGAGGTT  
17951 TCCAGACCTC AATTGTTGAC TTCTCTGCAA ATGTAGGCTC AACACCCCAT  
18001 GGAAGCTGGC AAAGCTTGGG GCTTTTACCT TCTGAAGCCA TGGCCTTAGC  
18051 TGTACCTTGG CCCTTATTAG TTAAAGCTGG AGCAGCTGGG TTGCAGGGCA  
18101 CCAAGTCCCT ATGGTGCATA CAGCAGGGGG GCCCTGGACC CAGCCCACAA  
18151 AACC AATTTT CCCTCCTAGG CTTCTGGGCC TGCATGAGT AGGGTTGCCA  
18201 CAAA AACTGTC TGACATGCCT TGGAGACATT TTCCCTATTG TCTTATTAAG  
18251 ATTTGGCTCA TAGTTACTTA TGCAAATTTT TGCAGCAGGC TTGAATTTCT  
18301 CCTCAGAAAA TGAGTTTTTC TTTTCTATGG CATCATCAGG TTGCAAATTT  
18351 TTAAA AACTTT TATGCTCTGC TTCCCTTTTA CAATTAAGTT CCAATTCCAA  
18401 ACCATATCTT TCTGGATACA TAAA AACTGAA TGCTTATAAC AGCACCCAAA  
18451 TCATATCCTG AACACTTTGC TTCTCAGAAA TATCTTCTAC CAGATACCCT  
18501 AAATTATCGC TCTCAAGTTC AAAGTACCAC AGATCTCTAG GGCAGGGGCA  
18551 AAATGCCACC AGTCTCTTTG CTAAAGCATA ACAAGAGTCA CCTTTGCTCC  
18601 AGTCCCAAC AAGTTCCTCA TCTCCATCTG AGACCACCTT AGCCTGGATT  
18651 TCATTGTCCA TATCATTATC AGCATGTTGG TCAAAGCCAT TCAACAAGTC  
18701 TCTAGGAAGT TTCAA AACTTT CCCACATCTT CCTATCTTTT TCTGAGGCCT  
18751 CCAA AACTGTT CCAACTTCTG CCTGTTACCC AGTTGCAAAG TTACTGCCAC  
18801 ATTTCTGGGT ATCTTTACAG CAGTGCCCCA CTCCTGGTAC CAATTTACCA  
18851 TATCCATTTA TTCTCATGCT GATAATAAAG ACATACCCAA GGCTGGGTAG  
18901 TTTATAAAGA AAAAAGAGGT TTAATTGACT CACAGTTCAG CATGGTTGGC  
18951 AAGGCCTCAG GAAACAGAAT CATGGTGGAA GGAAGCAAA CACATCCTCC  
19001 TTCACATGGT GGCAGGGAGA AGAATGAGCA AAACGGGGGA AAAACCCTTA  
19051 TAAAATCATC AGATCTCATG AGAACTCACT CTCTTGAGAA CAGCATGAGG  
19101 GTAACCATGT CCATGATTCC ATTACCTCCC AACGGGTTCC TCCCATGACA  
19151 CGTGAAGATT ATGGGAAC TAACAATTCA AGAGGAGATT TGGGTGGGGA  
19201 CACAGCCAAA CCATGTCAGT CATGATATGA GAAATTATCA AATTAAGATG  
19251 TAGGGAAGGT TTTTAAAAGA TTTGAGCAAC CACAAATGAC AGATATGTGC  
19301 TATAGTAGTG CAAAATACCA TTTTGCTCTT ATTA AAAATA TAATTGTTCT  
19351 TGATAATCTG AATTATAAAT GTCATGGATA ATTATGATGC ATTATGCTCT  
19401 CAGCAGCTAA AACTTCAAGC AAAATACACA CCTAGAGAGC AATCAGCCTT  
19451 AACAATAATT CTATAAATTT AATTTTCTTT ATTTCTGATA ATTACATTTT

Fig. 2 (cont'd 10)

16/124

19501 AGTTGACTTC ATATGTGATC TAAATACATT ACCATTATTT TGGACTTATG  
19551 ATGTAGCTCT TGAAGTACAT ATATGATGTA GCTCTTAAAG TACATATAGA  
19601 AGAGCAGATA AAGTATCAGT TCACCATTTT TTTGTAGTTT GTGCTTTCAT  
19651 GATGAATATT CTCATCAATG TACAGATTAT TTGCAGGAGC CTTTTAAATC  
19701 CATGTGTCCA TTTTATGAGA CTTAGCTTTT GTCTGTATAT AATGTGTTA  
19751 TTCAGTGTGC ATGGATTAAT TTGAGAGAGC ACAGGTATGG GTATCTTTAC  
19801 AGCAGTGCCC CACTCCTGGC ACCAATTTAC TGTATTAGTT TATTCTCATG  
19851 CTACTAATAA AGACTATATA TCACAATAAA CTGAGAACCA GCTGGTAAAT  
19901 GAGAGAAGTG TGGTCCACCT TTTCATTGTG GAGTTCTCAT TTTCTTAGC  
19951 TTATGCTGCT TATTCAACAC TATTTCTGCA TAATCTAATG CATTCACTAA  
20001 ATGAAGGTGC TGTGTTAGCC TCCACATGAT ATTAATACAG CCTATTTAAT  
20051 TTATCCTTCT TTAGATTAAA AATAAATAAG TAGTCATGTG CCACAGAATG  
20101 ACACTTCAGT CATTGGTCA TTGAAGGACC ACATCTATTA CTGTGGTCCA  
20151 ATAAGATTAT AATAACATAT TTTTCCTGTA CATTTTCATT GTTCTGATAT  
20201 GTTTTGATAC ATAAATGCTT ACCATCGTGT TAGAGTTGCC TGCAGTATTC  
20251 AGTACAGTAA CATGCTGTAC ACCTAGGAGC AACAGGCTAT ACCACATACC  
20301 TTAGGTGTAT AGTTAGGTTA TACCATCTAG GTTTGTATAA GTACACTCTA  
20351 TGATGTTCTC ACAATGAACA AAATCACCTA ATGATGCATT TCTCAAAACA  
20401 TGCCCTGTC ATTAATACAG TATGTAACAA TACAGTTAGT ACAATATGTA  
20451 ATACATGACT ATATTCAGAA TTTTAGCTAT TTCTCTTATA TTTCAAATGG  
20501 ATTTTCTTAT GCACTGTGTG GCACGGGCAT TTCATTTTAG TAACCACAGT  
20551 CTGGGAAAGG AGAAGTCTTT GAAGGATGTT GAGCAAGGTT ATGACATGGC  
20601 CAGATGTGAA TTTTGTATCA GTGACTCCAT GTTAGCAGAT AAAGTTGTAT  
20651 TGGGAAAGAT CAAAAGCATG AAGGCCAGAT AAGAGGATAC TGTATGTTAT  
20701 CATGGATGGA AATGTGAGGG ATGGCAGGAG AGATGCTATG ATTGAATGAA  
20751 TCTCAATATT CTTGGTGATC AAAGAATAAT GAGACTCATC CAATAAGACT  
20801 CTGTGAATGA TTGAATGTAG TTCCTAAGCT AGGAGGAAGA ATGAGGAATG  
20851 ATTTTCTGGT TCCTGACTAC AGCACAAGTT TTTGATTTTT AGAACAAAGA  
20901 ATAAATTTGT ACATGCTTTA TGATTCCTGG TTGAATTTTT AAGGATAAAA  
20951 AAGTCAGCTG TAATATTATT CTTTCCTGAT ACCATGCAGT ATTTGTATCA  
21001 GTGATCTTAT TCATTCCACA CACATTCTTC TTGAACCTGG ACACTGCTCT  
21051 AGACACTGAT TCTTTCCAAA TATCAGATAA GGTATTCTT ACGTAGACCC  
21101 TCAGTTCATA TAAATATGAT TTTCCCAAAA TGTGAAATAA GTGACTTTTC  
21151 ATAAGATATT TTTTAAAAGA ATGTCTTAAT AATAAATTGT GAATGTTGCA  
21201 TGGAAATGTA GGTGACTTGC ATTGTGCATC CTGTGTTTGA TTCACTGCTC

17/124

21251 TTGCATGTCT TGCCTTTAGC TGGGATGACA GCAGTTCAGT GAGCAGTGGT  
 21301 CTCAGTGACA CCCTTGATAA CATCAGCACT GATGACCTGA ACACCACATC  
 21351 CTCTGTCAGC TCTTACTCCA ACATCACCGT CCCCTCTAGG AAGAATACTC  
 21401 AGGTGAGAAT TACCACCTTT CTTTTTCCAG TGTTCCTGCC AGCTTTTTTC  
 21451 CCAAAATTAC TTAATATTAG ATTAAGGTAT AGCACAAGCC CTTAATCCAA  
 21501 AATTATTACA GAAACTGGAA AATGCAGAGA TAATAAGGAC TCCCTTTGCC  
 21551 ACTCCTGAAC CCTGAAGCAT CTTTCATCTT AGTCTTTCCT AAAGCCACAA  
 21601 CCCTTAGGAG GAGCAACAAT GTGCACTGCA GCCAATTTTG AATAAACAGA  
 21651 AGCAGCTTAT ATATATATAT ATATATATAT ATATATATAT ATATATGATA  
 21701 TACATTACAT ATTTATATAT ATGTAATATA TGTGCCATAT AGCCTGGTGG  
 21751 TATAGTTATC TATACAAATA TATTTATTTA TTGTTAATAT ATAGAGTATA  
 21801 TAAATATCTA TTTATATAAT AGATATTTAT ATATATTAAA TATCTATTTA  
 21851 TATAATAGAT ATTTATATAT ATTAAATATA TAAAAATATA TAACATATAA  
 21901 TAGATATATA TTTTATATAT TATATAAATA TATATTTATA TATTTAATAT  
 21951 ATTAATGATG AATTACTATA TTTGTATAGA TAACTACACC ACCAAGCTAT  
 22001 ATGGTGTGTA TATATTAATA TATAATGTAT AATTCCTATAT TAATATAATA  
 22051 GTAACATATC AATACTTAAT ATAATATATA TTCAATTGAT TACAATCTAA  
 22101 TTCAGAAAGA TTTATGTTGC CATATCTCTC CTTACAAATAT CGATATGTTT  
 22151 GTTTAAAAAT CCAGCAATTA TTTTCATAGT CTAATTTTAG ATAGTTCCTG  
 22201 ATTAATTTTA TATGATCTCT GAAATATATC ACTGGATCTG TTGTGAATGA  
 22251 TAAATCAAAA ATGAAAAATG GACATTACAT CATTAAGTTC TAGCTTGTCT  
 22301 TACTACTTCT TATGACATTT GATATAGAAA ATTTCTACCT TTCTGTAGCG  
 22351 TTTAATTTGGT GTTTTCTGCA TGTATTTATT CTGAAATTCT CTAATATCTG  
 22401 CAAGTGGGAA TTATGTGGCT AAAATTAATA AAATGTAAGT GAAGGTAAAT  
 22451 CAAAATAGAA TCTTTGGATT TATCCAGTTA TCTGAAAGTA CATTTTCATTG  
 22501 CCTTAATTCA CACTTTATAA ATTTTCTAC ATAAAGTTT TCTGTAATAT  
 22551 TTGTCTTTAT AGCTGAGGAC AGATTCAGAG AAACGCTCCA CCACAGACGA  
 22601 GACCTGGGAT AGTCCTGAGG AACTGAAAAA ACCAGAAGAA GATTTTGACA  
 22651 GCCATGGGGA TGCTGGTGGC AAGTGAAGA CTGTGTCCCTC TGGACTTCCT  
 22701 GAAGACCCCG AGAAGGCAGG GCAGAAAGCT TCCCTGTCTG TTTCACAGAC  
 22751 AGGTTCCCTGG AGAAGAGGCA TGTCTGCCCCA AGGAGGGGCG CCATCTAGGC  
 22801 AGAAAGCTGG AACAAAGTGA CTCAAAACAC CCGGTAGGCT TGTGCTTTCG  
 22851 CAGCTGTTAT GCAAAAGTGC TTTACTTTAT TGTTTCCATT CAATCTTTGT  
 22901 TTTCTCTAAC AATAGCATTT CTAAAATACC AAATTCTTAT CCATATTAAA  
 22951 CATGGAGTCA AATAGTTAAA TAGTTTTTCT GTCTACGTTT CACAAACTCG  
 23001 TCATAGAAGC CCAAGTAGGG CCTATATCTA GGCATTCTCT GGAAAGCCTC

Fig. 2 (cont'd 12)

18/124

23051 CTCATAAACT AGGGGTACTG GATGCCCTTAC CTTGCCAGAG TTATTTTCAGG  
 23101 TAATGGGGAA ATAAGATTAG GTTGCTAAAG CAACAGTTAA GTTTTTTTGT  
 23151 TTTTGTTCCTG CGTCTTAAT GAAAGTTTGG AATTTTTTACA CTAAATATGC  
 23201 CACTGAATTG CACTACAGAC TCTGAGAGGA ACAAGCAATG AACTAATCA  
 23251 ATTGGAATGC TGGAGATTG AAATATTGTC TGTGTATTAG ACTTCATGAA  
 23301 AGAAGAGAAT GAAATAGTTC TTCAAAATTG TGCCATACTT TTTTAAAAA  
 23351 GACTCTCCCC GTATTTTTAA AATAATGCCT AATTATAAAT AGTGCCACCT  
 23401 GAAGCACTAA TTAACAGGGT ACTCCAAATA TAATCATCTC ACAGATATTC  
 23451 AAATGAATTC TTTTCTAGT AATTAGCTTG ATAGGGTTAA GTGTTACCTT  
 23501 TTTAAAAAGA GTTGCAAAAT ATAAGACATT AACAAATAGC AAAACATATG  
 23551 TTTTCATTTT ATCTCTCCA TCTCTCATAA TGTTTCTTCT GACAGCCAAA  
 23601 TTTTTGTAGC TATGCACTCA GTCCTCTCAA TATATGAGAT TTTTGATCTA  
 23651 AGCCAATACA TTTAGGAAGG GAAATAATAT AAAGAAGCAT TCACATTTTA  
 23701 CACATTGTTT CACGAAGTGT GGTGATATCA AACTCTACAG GCACATATAT  
 23751 TTGTGTATTT CTCCTTAATT AGGGAAAACC GATGATGCCA AAGCTTCTGA  
 23801 GAAAGGAAAA GCTCCCCTAA AAGGATCATC TCTACAAAGA TCTCCTTCAG  
 23851 ATGCAGGAAA AAGCAGTGGA GATGAAGGGA AAAAGCCCCC CTCAGGCATT  
 23901 GGAAGATCGA CTGCCACCAG CTCCTTTGGC TTTAAGAAAC CAAGTGGAGT  
 23951 AGGGTCATCT GCCATGATCA CCAGCAGTGG AGCAACCATA ACAAGTGGCT  
 24001 CTGCAACACT GGGTAAAATT CCAAAATCTG CTGCCATTGG CGGGAAGTCA  
 24051 AATGCAGGGA GAAAAACCAG TTTGGACGGT TCACAGAATC AGGATGATGT  
 24101 TGTGCTGCAT GTTAGCTCAA AGACTACCTT ACAATATCGC AGCTTGCCCC  
 24151 GCCCTTCAAA ATCCAGCACC AGTGGCATTTC CTGGCCGAGG AGGCCACAGA  
 24201 TCCAGTACCA GCAGTATTGA TTCCAACGTC AGCAGCAAGT CTGCTGGGGC  
 24251 CACCACCTCG AAACCTGAGAG AACCAACTAA AATTGGGTCA GGGCGCTCGA  
 24301 GTCTGTGAC CGTCAACCAA ACAGACAAGG AAAAGGAAAA AGTAGCAGTC  
 24351 TCAGATTGAG AAAGTGTTC TTTGTCAGGT TCCCCCAAAT CCAGCCCCAC  
 24401 CTCTGCCAGC GCCTGTGGTG CACAAGGTCT CAGGCAGCCA GGATCCAAGT  
 24451 ATCCAGATAT TGCCTCACCC ACATTTGAA GGTAAGGATG TATAAAATGA  
 24501 TGCTGAAAAA ATATAAAGGA TAAATATGTG TTAGACACAT ACATTACATA  
 24551 TAAATGTGTG TATATATATA TTTTAAATAT GTATAAGGTA TATAATATAT  
 24601 ATATCTTAGA ATTCTTTAAA GTACACAGTG AGCTCTATGA AGCTTATCAT  
 24651 ATAAACAGCT AGCAAAAAAA ATAGTTCTCA TTTTGAGAAA CAGTCAAAC  
 24701 TCAAAGTTTC ACTGTCATTG TGATACTAGC AACACAAACA TCTAAGAGAC  
 24751 TTAAAAGCTG ATGGTAATAC CTAAGTGTAG TGATAAGGCA AAGTAATAGC

Fig. 2 (cont'd 13)

19/124

24801 TTGTAAATTT TCTATAGATT TCCATTCCCTC CTTTTCACAT TAAAAATTAA  
24851 AACCAAATAG GTTTTCATGA CTTTGGGCAT TCATTTCAG TGTCAATTTTC  
24901 TTGCTGGCTC TTAATGAGTT GGTGATCATA AATGTAGATG AAGTTGTTTT  
24951 CCTTGTAACA GATTCCATTG GACAGATTTA TACAGTGTC AATCTTGACA  
25001 CATTAAAGAC AATCAAGATA TGACATAATT TGAAACTATT CCAGTGTTTG  
25051 GTACAGTATC ACAACTGAAG AGTGGGCTAA GCTTCTAAC TCTTCATCTG  
25101 CTTTCTTTGA CATGACTCTG GTAAGGATCA TGACTTGGTT TCTGTTCCTG  
25151 GATTGTTTTT GGTGTTAAAT ATGTGAAGTT CTGCTCTAAG ATATCACTGT  
25201 TTTTAAATAC CCATGTGTTT TTAAGTGGTA GGAAAATAAA TGCAGTTAAA  
25251 AATTGGGGAC AAATATCTAA ACCTCTCTGA GTCTGTTTTT TCATCTGCAA  
25301 AATGGTAGAG TGTGGTTTAT AGTTCATTAT GGGTTCAATA TTTTAAATGT  
25351 TTGTTTTTAT TCTGTTGACT AAACCCAGAA CTTTGATATC TTGGAAAGGA  
25401 AAGATTTTGA AACATTTATT TTACAATAAA GCAATTTTCA ATACCTGATT  
25451 GTTTGAAAAA CCTAAAGGCT TTATTCCTCC GTAGTAATAT TAATGCTGCA  
25501 GAACTGTCTT TTTAAAATAC TGATTCTCAT TGGGAAGAAT GAATTATGGC  
25551 GTATAGGGAG AGTAAATATT TCTGTTTCTT AAGTAAAAGC CAATAGTGCC  
25601 CTCCTGTGGC CCATTACCTA TGAAACAATT TCTCATATTC GTCATAAAAT  
25651 ATTTCACTGT AGGAAATATG GATTTTATTG CAACTCAATT AGTAATCATT  
25701 ATGCCATTAC TTCATATCAT TGTATTTCCA TATTTACATA AATTTGATTC  
25751 TACCATCTGC TTCATTTACA AAATAAAAT GTTTTCTGAA CTAAACTCCA  
25801 AAATCTAACA GCACCAGCTC TGTTTCAAAT CACTATTAAT AAATGTATTT  
25851 GAATAGCACT GGCAACTGAC ATAAAACCTT TTGGCCTCTG CTGGGGAAAA  
25901 TACAGACAAA CTGACTTGTT GCCGACAATA TCAATATTGT TTCCAACCAA  
25951 CTGCTCCCTG ACAGTGACTC AGACCACCAG ATACTCAACA CAACTCCCTA  
26001 AACTTGCTTT AAGCGTTCCA TCTAGATTTT GAATAAACTG TTTAAAAATT  
26051 TAAAAATAAA AAAAAAGAG AAGAGCTCAT TTAAGTGTG TCTATCGAAT  
26101 GCGTAGAAGT TGTTTCATTA TAATGGTTCT GTAAATAGGT AACAGCAAGT  
26151 ATGGTCAAAC TACTGACTTT GAGTGAAAGT CTCATGATCA CTAAATTAT  
26201 GAAAACCAGG GGTTTTCATG TTTGACTTAC TTTTGTCCA CCCACTTCCC  
26251 CTCTTCCCT AGTAGCAGCT CAGTACTGAC CTACCCTTAT ATGAGAGATT  
26301 TTCTGCACTT GATAAAGAAG TCCAAGCTTA TAAAAGTTCA TTAACATAGA  
26351 GACAGGAAGT GCTTTGTAGT TCAGTACATC AAAGCACACT TGGCTCTGTG  
26401 TACTGTAACC CGAAATATTA AATGTGGATA TTAGCTTCTT GGAACAACCTG  
26451 AAGTTGTTAT TTGTTTTTCT TTTAGGTTGT TTGGTGCCAA GGCAGGTGGC  
26501 AAATCTGCCT CTGCACCTAA TACTGAGGGT GTGAAATCTT CCTCAGTAAT  
26551 GCCCAGCCCT AGTACCACAT TAGCGCGGCA AGGCAGTCTG GAGTCACCGT

Fig. 2 (cont'd 14)

20/124

26601 CGTCCGGTAC GGGCAGCATG GGCAGTGCTG GTGGGCTAAG CGGCAGCAGC  
26651 AGCCCTCTCT TCAATAAACC CTCAGACTTA ACTACAGATG TTATAAGCTT  
26701 AAGTCACTCG TTGGCCTCCA GCCCAGCATC GGTTCACCTCT TTCACATCAG  
26751 GTGGTCTCGT GTGGGCTGCC AATATGAGCA GTTCCTCTGC AGGCAGCAAG  
26801 GATACTCCGA GCTACCAGTC CATGACTAGC CTCCACACGA GCTCTGAGTC  
26851 CATTGACCTC CCCCTCAGCC ATCATGGCTC CTTGTCTGGA CTGACCACAG  
26901 GCACTCACGA GGTCCAGAGC CTGCTCATGA GAACGGGTAG TGTGAGATCT  
26951 ACTCTCTCAG AAAGGTGAGC TTTCTGGAG GCATTGATAA CATCTTCCCC  
27001 CTCTTCCCTG CACTATGCCT AACCCCCACC CCATTAAATT CCCTTGATTT  
27051 CACTGTGAGT GCCCCGGTGC AAAAAGATGT AAGACTGATG AAACCGGGCC  
27101 TTTCATTTGC TCTCATTACC AAATTTACAG AGGAATAGAA TCATTAAAGG  
27151 TAGGGTGAGT GGATAATTTT GTTAATATGA ATGCATACAT TTATACCCAG  
27201 TAGGCAATGT GAATAAAATT CAAGGAATGT ATTTAGATAT TGAATGAGGT  
27251 CTCCTGAAGA CATTTTAATG ATTTGGCTTA AGCTTCAGAA CAACACTAGC  
27301 TCCTTATGAT GACTTAAGCA TTTTGAAAGA CCAAATTGAA ATTATTCTAT  
27351 AGTTATGCTC AGAGCAATAT GTTAAATTTG TTCCATTTGT ACTTCTATGA  
27401 AAAAATAGCA GATGGATTGC TGGGAAATCC TAGTTGGCCT GGTAAAAAAA  
27451 AAAAAAAAAA TCAATTGTCA GCCATGAATC ATTAGAGAAA ATTATAGTGT  
27501 CAGTGCCATT TTCAATAGAC TGCTTAAAAA GTAATCATAT TACAAAGTGT  
27551 TTCTCATGG CTTTATATAT ATATATAAAC TTAAAGTAGA GGACATAGCA  
27601 AGGCATTTCT TACCTAATAT GCTTACTGTG AAGCATCCCT TTTGAGCAA  
27651 ATCACTCTAA ATTTTCTCCT CAAAGTGATC CTCTCTTGAT TATACTGTAC  
27701 TGACTCTTAC CACCAGGAAA ATGTCTTAAA ACCACTTCTT TTTCTGATA  
27751 AATGCAATGC TATTGTCTC TTGACATAAG TAAAGCTTTA AACATGGTCT  
27801 TGGCCACATG TGGAAAGAAA TACTGGTCAC GTAAAATACC TGATATATCT  
27851 TTCTATGTCT TCCCCTGTTT TTTTATTTT TTTTATTTT TTATTTTTTA  
27901 ACTCTGATAT TGATGATGGC ATTTATTTTC TAGACCTTCA GCCTTACTCC  
27951 CGGAATGATA TTTTAAACA TCAATTAAAG CCCTTAGCTA GACACTCTCT  
28001 GCATTACGCC AGTTTCCCCT TAATGTAGGA TGTCCCAATT TGAAATTCCC  
28051 CATTTTCTCT TGACTTTGTA AAATACAAA CCCAGAGCAA AACATTGCTT  
28101 CTTTCCCTCT TTACTTCCTA CTTGCCTAAC AATGAGACAG GGACAGCCGT  
28151 GCAAATGGGG CTTTCCGATG ATAAAGTAAT TTAAACACTA ACTAAAATAT  
28201 TGGTGTTC TATGGTGGGC TGCTAATTAC AAAATACATT TTTCTCCTA  
28251 AAGAAAAAAA CTGGGCCAAG GCAAACAGCT CAGTGATAGC AAATAAAATG  
28301 TAACCATTTC CCTATGGTTT TGCTGTTATA TGCTATTATA GACAGCATAC

21/124

28351 GTAAAGACCA GTAAGGGTTC ATTTTTCAC CTAAATGTC GGGCTTCCTG  
 28401 TAAATCTTT GATTCTAGTT TCAGCACTTC TAAGGTAAAT GGGCATCTTC  
 28451 ACATGTCATT TATAAACTT CTAATGAATG AATTATATTA AAATAGATAA  
 28501 ACAACCTATA GTTTTAATGA ATGTATCCTA GATTGTATGC TCATATGTAA  
 28551 GGATTCTAAA TATCAACTTG ATAACCAAAC CAAACATAGT GCAAATAGGT  
 28601 TATCATTTAT TAACCACAAC CACCTTCCAC AAAACTGGTC ATTTTAAAT  
 28651 TATTAAGATA ATCTGCAACA AGTGGCCAT TTAGCCATCA GCCTATTTCT  
 28701 TCAGCATTTA GACATTAATC CCAGATTCAG AAATAAGTC AAGTAACTAT  
 28751 TTATAACCAA GTAACATTCA AATCAAACT AGATGAAAGA TTGGTTAGTT  
 28801 GCATAGCTAT AACCAAAATG CAGTTTAAAT ATTTTACTCT AATCTATATT  
 28851 TTAAGTGAAG TCAATAAAAT TTTCATATG GAAATACACT AGAAAATATG  
 28901 CAATTTCTTA TTCTTTTAA GCAGATTTAT TTATTGTACA TGTTCAGTCT  
 28951 TTGAAATAGG CCAATTTTAT TTATGTTATG TTATGTTATT TATTTGTTTT  
 29001 GAAATGGAGC CTCCTCTGT CGCTCAGGCT GGAGGGCAGT GGTGCCATCT  
 29051 CAGCTCATTTG CGTCTCTGC TACCCGAGTT CAAGCAATTC TCATGCCTCA  
 29101 GCCACCTGAG TAGCTGGGGT TATAGGAGCG GACCACCATG CTGGGCTAAT  
 29151 TTTTGTATTT TTTGTAGAGA TGACGTTTCA CCATGTTGGC CAGGCTGGTC  
 29201 TCGAACTCCT GACTTCAAGC GATCTACCCT CCTTGGCCTC CCAAAGTGTG  
 29251 GGGATTACAG GTGTGAGCCG TGGCACCAGC CTGAAATAGG CCAATTTTAA  
 29301 AAATGGGAGT ATTCCTACAT TAAATGGCC AAATAAGAC TTTTCTAAA  
 29351 ATAACTTTA AACTAATTTT GGATAAATAT GTTTTGCCTT TGAGCCTTAA  
 29401 TAAATGCAT TAATGAATAT TAAGCTGTAA AAAGTACATG TTAATACAT  
 29451 AGCTATAGTG TATAATATTA ATATTAATTA GTGCCTTCCA GTAAATTAAT  
 29501 AGATTAAAT AAATTTTAAAT ATAAGACACT GAGCTTTTTG TTTTCTTGAC  
 29551 AATAGAAGTG CAAGCAATAG CAAATGCTC TAATCCTTTC ACGTACATTT  
 29601 AAGAAAGTTT ATGACCTATT GAAGAGAAAA GTAGATCTAG TGGGTGATAC  
 29651 TGGCTTCATT ATGGTTAATT AATTGATCAG TAGAATGTCA GAAATGCTAA  
 29701 GAAAACCAA GAACTACACC AGAGAGAAAA TGTGTTAATG TAAATTTTAA  
 29751 GGCAAGTTAA TTAGCGATAT ATAATAAGA TGTATATAAG TTCATGATTT  
 29801 ACCTGTTTGT CTACAATTTT AGATGATTTT TTGATACTCA TATTTAAATC  
 29851 GGTAGCTTTT CCTATAGATT TTAATTTTTG TTTAAATTCC TCTTCGTAA  
 29901 ATTAAATAAA ATAATAAAAT AACTTTTTTA ACAGTTTTCT CTTCTGCAGC  
 29951 TGCTCTAGGT CATTTGGTGGC CATTGAGCCA TAACTAGTCT ATATTTGTTT  
 30001 TGGGTTTTGT TTCATGTGTC TGAATCAACT AAATTTTTAA ATAATTTGTA  
 30051 GTAACCAACT TTGCAATTC TGGGTTTGTC TTTAAATGTC AGATCTGGCA  
 30101 ACGCTGCCTT GACATTTCTG CCTAGAACT ATTGGCTCTA GGCAGTCAGT

Fig. 2 ((cont'd 16)

22/ 124

30151 GTCTGTCTGC TTCAGACTGT TGA CTGAAAT CCCCATTCGT TTTCATGCCC  
 30201 TATCTGGCCC TTGCTGGCAT ATGAGTTTGC AACCTTTGGT GATTTGCAGA  
 30251 AATTGTCTAT GTTAGAAAAT CATTAATATC TAGATTCAA CATATTTCTA  
 30301 AATAAAGCTT TAAATTATTA TGGTAACTTT AAATGTATTT ATTCTAATTT  
 30351 TTTTCATTAA ATTGCTCTTC ATCATATAAA TATATAATTT TTATACAACT  
 30401 GGATGAGTTT GGCAGAAGAA TACCAACTTT TCATATTCTT TGTGGCATT  
 30451 AACTTTAACT TGTACACATG GAAATAAATA ATCCTTAAAA TGACTTATGA  
 30501 CCACATAAAT GCCTTAGCAC ATGTGGTTCA TATTTGGAGA TTTCTCATAT  
 30551 TTGTTCAATA TAATTTATTT TGTTTGTTTA TCCACAGTAC TTAAGAAAAC  
 30601 TTCTATAGTC AACATATATA CTGTAAGTGG CCTCTACACA GTATAAGCAA  
 30651 TTACCTTACA TGGCTATTAC CGATAAAGTT AAAGTTGTAT AAAGCCTTTG  
 30701 GATGCTTTTG ATTTTCAGTGC TAAATAATGG AGTACACATA GAAGAAAACA  
 30751 TTTTAGCTTT GGT TTGAGTG ATCAAATTT AGGTCAGCCT TTTTACATTC  
 30801 ATGTTATATC ATCCCCATTA TCGGTATCCT GTGTATTTAA TTTTGATCAT  
 30851 TTGATGTCCT AAAGGAAGAA AGCTATAATT CTGCAATTTT AATTAATTTT  
 30901 ACACTTTGCT TATCCACATG CCAGAGATTA TAAAAGAAAT CCCTAAACTT  
 30951 GTCCCACTTA GTTGTGATA TCCTCTTCCT GTATTTT TAG AGAGGCCATT  
 31001 TCTTATTTTC TCTAGACATA GCTTTTCATT CCTTCTTGTT ACCAATTGTG  
 31051 AATTCCTTAA AATAGAGATG ATAAAATTTA TAGCCTTTTA AATACCTAAT  
 31101 TTATGATTTT TAAAAGATGG TATAGCTTAA TTTCATTAA ATATTCAAAT  
 31151 AAATGATACT AGAATCAATT AAGTTTAAAG CAAACATTCA TATATCTTTC  
 31201 TTCACATGTG TAAATGGGAA ATAAACATGC CTTTTTATTA AAAATAATTT  
 31251 GAAGACAAAA GATAAGTATT AAACAACGTT TTATACCATC TCTGTCAATT  
 31301 GGAAGTTGTC ACTCTAACTT AGCCAGAGCA GATCTATCTC ATTTTGCATG  
 31351 TGATATCATA GCAAAAGTCT AATCAGTTGC ATAGGGAAGG AAAAATAAG  
 31401 ATAGTATTTA ATCAATAGGA TTCAGAGGAA AATTATGCTA ATGTGATTTA  
 31451 ATCTATTTTC TAGTAATCCT ATCACTAAAC TGTCATTGAA TTGTACTGCA  
 31501 TTAGAAAGGA ACTCAAATAT GTGTGACGGC AATGGACATC TTGTCACCTT  
 31551 TAGTTGGCCT TTTTCAATGA GTTAAGCATT ATATGTGTGT TACCAAAAAA  
 31601 TTATTTT TTA TAGTTCAGAG AACCATTTTT GTTGGATGTG TAATTTGGAA  
 31651 GTTTTGT TTA CAT TATGTCC TTAGGGGTTT TCTTTGTTTT AACAGCATGC  
 31701 AGCTTGACAG AAATACACTA CCAAAAAGG GACTAAGGTA TATATTCCTC  
 31751 TCAGCACAAT TGCTACCTCT CTGTTGTTAT GTAACTTTG TGTGCTGTCT  
 31801 CTCTTCCTTC TTTGTTTGTG TGCAATGTAG CACATGACAT TGAGGACGAA  
 31851 ATCACTTTTA ATTTTGATGG TTTCTCTGGC CCGAACAGTT GGTGAGATAG

Fig. 2 (cont'd 17)



23/124

31901 CCCCTTAGGT AGAGATACTA GTAGAGATTG AGGCTGTCTC TCAAATTAAA  
 31951 TAAATTCCAA TGTGAATATC ACTATTTTGA AGAAATAATA CTAAACAAAC  
 32001 AAACAAACAA AACAAAAACA AACAAACAAA AAACCTGTCC CAGGCATTAC  
 32051 TTTTTTGGGG GCAGCAACTT TGGTAGAATG CAGAACTCAC TTCAACAAAT  
 32101 TAAAAATAAA TTAACCTCTC TAACTTTTGC CTATTAGAGT CATATGCATG  
 32151 CAAATATTCA AAACCCATGC AGTCTACAGA TGTGGGCAGT TAATGTTGAT  
 32201 AGGTTGAAGG ATGCTACAAT CTGAATCAAA GAAACATAT TTTCATCATC  
 32251 ACAGGACAAA TGCTGTAATT AAGGTGTGAT TTTTATAGAA TCCTTTTGAT  
 32301 AAAATCTCAA AATTGTTTTA ATTTCTATTT TGCAGGGGTA CTGCTATCAG  
 32351 ATCAATTTAA ATCTGAATTA ATCTAATATC ATTTAATAAT CTCAAAATAA  
 32401 TTATTCCATC CATAATAAAA AATAAAATAA AAATTTAACT TATGGCCATC  
 32451 TTTTACTGTG TACTTTTATC TGAGGAAGAG ATAGAATGAT CTAATAATAG  
 32501 AGGTATAACA CTGTATGTGT ATGAAAAGTT GGCTAATTTT GGTGCTAAGA  
 32551 ATTTACTTAC AAAAAGAAAA AGAATATACT TAGTTTGGTG AAACACTGAA  
 32601 TAATGGCGAA ACTAGGTCTT TCTCCATTAT TTTTTTCTC TCCAATTTT  
 32651 CAGCAATAGC AAATAGCTGG CAATTATTCC ATGTTAATAT TTGATCCAG  
 32701 AAATTTATGT TCCAGTAAAG CGAGCACATC TCCCTCCTTA TTTTGTAAT  
 32751 CTAGGCATGA TGCAAGTGG CAGTTTAACA AAAGAACTGT TTTTCCTTTA  
 32801 AAAAAAAAAA AAAAACAAAA GCTGCCAATA TGTATTCCAT TTCCCTATGC  
 32851 CTTCTGTGAC CATCCTTCAT TTCCCTTGGC CCTGGCCCAC CACTGTCCCTC  
 32901 CATTTGTAGT CCATGTTTTC ACCCTCTTTA CATCCTTTCT TGCCCTGTGC  
 32951 TTTTGAGTTC TCAATTAAC TGGCTGTCTG CTCATTGCTT ATGATTTCCTA  
 33001 ACTGCATATC TGATAGAAGC ATAATTTTCT CCTCAAAACC CTTTATCTTA  
 33051 TTTTTTTTCC CTATGTGATT CAAACAGATG GCGTAAGATC ATCTGGAAGA  
 33101 ACTGAGCAAT TATAATTAGA TTCAATCTGT TTGAAATTGT TCATTCTGAA  
 33151 TAGTAACCTC CTCTGAATTG TTTTCCTGTC CTGGCATTGC CTTGCCCTTG  
 33201 TAGATGTGCT TAAGTGTGCT AGCTGTGCTG TTTTGCAGAT ATACCCCATC  
 33251 ATCTCGGCAG GCCAACCAAG AAGAGGGCAA AGAGTG GTTG CGTTCTCAT  
 33301 CTACTGGAGG GCTTCAGGAC ACTGGCAACC AGTCACCTCT GGTTCCTCT  
 33351 TCTGCCATGT CATCTTCTGC AGCTGGAAAA TACCACTTTT CTAACCTGGG  
 33401 TAAATATATC TAAATATATG ATTTTGTTTT GTTCTTTTCA CCACCCACTC  
 33451 TCACAGAAAC CCTGGAATCT CTCCATAACA CAACACGTTT TCATTTAAAG  
 33501 GGAGGGATAA AAGCACTTTA ACAGTACCTT TCATTGTGTG CATTGTTTAC  
 33551 TCTTCACAGA AAAATCTCCA AACATTATGC TATTTATTGC TCATGACAAA  
 33601 TGCTTAACAT AGATTAATAC TGTGGTTGTT TTCTAGTCTA GGCTCCAGAG  
 33651 GCTCAGAAAG TTCACCTGAC TTGAAAAAGT CTTACCATTA CTAAGGGTTC

Fig. 2 (cont'd 18)

24/ 124

33701 AAGGCAGTAA CCAGTTCAGA ACATCTGACT TTAATCCCAG GGGCCTTTCC  
33751 ATTCCATTTA AGAATCCTCT TAAAAACAG GAAGGCATCT CCTTATTTAT  
33801 TTGTCTGAAA TATTAAAACA TCCTTAAAC AAAATTAGTA ATCTTTTGTA  
33851 GAAAATAGAA ACAATTAGGA AGAAAAAAT ATGTAATTCC ATGACTCAAA  
33901 GTTAACTTCT TTTAACACTG TTAAAGTTAA AACTCCTTAA AATTCATACA  
33951 AGAATTTCTG TTAAGACAAT ACTCTGAACA TTTTCAAATA GATACAATGA  
34001 AAAATAAATT ACCAACTTAG TCATTGGGT ACTTTGTATT TAACATCATT  
34051 TGTATGAAAT ATAAAAATCAT TTGCATAAAA TTTCATTAAA AGCACTCTGA  
34101 GTAACAAAAT AATTAAAGAA AACTAAACAT GCCAGATACC ATTTAATAGA  
34151 TTCAATGACT TTAAAAATAT ATTTATTTTC TATAAAGTCA CATATAAAGT  
34201 ATTTTCATTA TTTTATGTTT AAATATTTTT ATTATTAGTT TATCAGAAAA  
34251 ACTTGTACAT AAAGATGAGT ATTGATACAT AATCTTATTA GAGCCAGAGA  
34301 CGATCATTC CTTCTAGAAA ACACATCTCT GAATTTAGGA CGGAGGACAA  
34351 TGAAACAAGA AATTTCACTT TATAATTTAC CTTTGTCAAA CTATCCCAGA  
34401 GCACATCAAT TCCATCATGA AAGTACTCTT TTGACATTAT ATAAAAAATT  
34451 AGTAATAGAA AACACACAAT CCAAAACCTT ATATTTTCTA AACTTCAAGT  
34501 TAATCATCAA CTTCTCTTAG ATTTTGAAG ACCTGAAAAT AAACATAATT  
34551 TCAAATAACA GAACCAAAC ACCATATACA TTTGTAATGA GGCACAACAG  
34601 TCAATTTTGA GCCTTGTTAT TTCCAGGTTT TAGCTGAATA ATCTTCACTG  
34651 CTTTCTTAGC TTTTGGCCAG TCTAGTTTGG GGAATTTTT GCCTTACTGG  
34701 GCCTAAACAG AGTGTAATAT TAAAAATATG TAATAAGCCA TACTGAGAAT  
34751 AAGATAAATG CAGGTTTCTA ACTCCTTAGG GACACAAGTG GGGACAACAC  
34801 ATTCCATGAA CACAGGTGAA TGAATGCCCC TAGTTTCTCT GAGTTGGACA  
34851 ATTCATGCG ATCATTTTTT TCTCTGAGGC CAAAGTCTCT GGTGTTGATCT  
34901 TCCTAGCAGC TTCCAGAACA GAAAGTGAGT TTACTTTGTC TCCATATTCT  
34951 TTTTCTCCAT GCTCGGGAAT CCCCTGCTTT CCTGATCCCA CCACAAAAAC  
35001 TCCCCTGAGG ATGAAGCCTT GGCTTTCCAG GCTTCCAGGG AAGCCTCGAT  
35051 TCCTGGCTGG AGGTAGTTGT ACCACACTCC CAGAGGGCTA AATCCCATAA  
35101 ACATCATCTT CTGTCTTTGT AGATCATAGA ACTTTTTATT ATCATCCAGG  
35151 AAGATTTCTC TTTTGAAACA AGGCTGGAAA AACTTTATGT CAGTCTGAC  
35201 CTGCTCTTTA ATGACTGCGT AGAGGGAGAT GCCCAGCTTA TCCAACCTGG  
35251 GTTGACAGAG GGACAGATCT GCAGCCCTC TTGCCAGAGA AAACATCCTG  
35301 GCACAGCCAC AATCACAAC CCATTCTTCT CCCGATAGCT CCTTTGCTTT  
35351 GAAACTCATT GGTACTTCT CCAGTGTTT CAGGTCTATA TTCTCCAGGT  
35401 ACTCCAGCAC CTCTTTCCAG GGCTTGGACA AAAATACATC TGTGTTGGCC

Fig. 2 (cont'd 19)

25/124

35451 AGCATCAGTG CCAAGGCAGC AGCCTCCAAG GGCTCCTGCA CCCATGGACC  
 35501 ACATCCACAC AGAGAAGCAC CTTGGGTCCT CAAGTGCCTC CCTCTTCTTC  
 35551 CCTTCTCCCA AACCTGAAGC CCAGACACTA AGGGGTCAAA CCCTCCTGGG  
 35601 CCCTGAGGGT TCCAAGGGCC TCATTACTTT TTCTTTTTTT CACTGGAAAA  
 35651 AAAATTCTAA TCATGCACCT ACAGAAGATT GACATTTTTC AGTAAGTTGG  
 35701 ACTTTCAGC TTTTCAGCCAG GACAAGACTC AAGGCTATGT CTTTCTATT  
 35751 GCAACCCTTC CCACTATATT GAGTAGGGCT TTTAGCAATT GAAAACAATT  
 35801 ATTTTGGTCA TGGTTTCATA TAAGCTAATG ATTTTCATATC AAACACCAAG  
 35851 TTTTGTGTTT CTAACCTATA TAGTGATAAG AGAATTTACC TATAATGCCA  
 35901 AAGAATGTAT AGCTTTTATT TGCTTTAAGA TGCAGTTGAT TTTTAAAAA  
 35951 AGCGAAAAGC CTAACACTTT AACTTCAAAA AATGAATTTA AAATGTTTGT  
 36001 GTAGGTCATA GGAATATGAA AAAATTTTAT ACAACATCTA AAACACACCC  
 36051 AAATCACCTA AAGTGCTATA AGCTTGCTAA GTACTTCATG TCTCCTATCA  
 36101 ATTCTTTCAT TAATTGACGT TAATTTGATT AGTTGACTCC TTCTTCTATT  
 36151 TTTCCTCACC ATTATTATTC TGATTAAATC CACCTTCATT ATTCCTTAGG  
 36201 AACAAAAAGA CTCACCACTT AACTATGTCT GACATTGGTG AAGTCGTTTA  
 36251 AACTTAATTT TCTTATCTCT TGAATGGATA CATAATACCT AGGTTATATT  
 36301 GTAAAGAATG ACGGATATAG TGTATGTAAA GATGGAGAAG TGTGTAAGAC  
 36351 TTGACAGATT CTGCCAAATC ATTATTTTCA CTGGAAAGCA TGTCTTACAC  
 36401 GATCATAGAG TAGCATTCAT CAGATATGCC TGAGCTTTGT CTACATTTAA  
 36451 TTGAGTAGTA ATTCGCAACA CAGTAACCAC AGGATTTTAT GTAAAAGACA  
 36501 TTCACAGATT GTGT'TTTTGA AAGATTGTAT TTTTGAAGTA CAAAACATATG  
 36551 ACATTGTTAT CAAGGACTCA TTTACCACAA ATATCAAATA TTTGTGCAAA  
 36601 GATAAGTTTA TGCTAAGATT TGCATAAATT AAAGTTAACA TGGCAACTGA  
 36651 AGCTAACATG TCCATGGTCA CAATGTGTTA AAAAATGAAT GGTCTGTAG  
 36701 CACACTTGGG AATGTATTTT ATTACATAGT TTTCAGAGTT AAAACACAAT  
 36751 TAATAAATGA AATGTGAATT ATACTTTTAC TGACAACAAA GCTCTCTGTA  
 36801 GAGCTTTAAT GTTCTAATGA ATTAGAAAAC CACTGATCAA ATACATCCCT  
 36851 TACATTTTCAT TGCTATAGAA ACCAAGTCTG AAAGGTAAAG TTTACCTTTC  
 36901 TAGGATGTGG GTTTCCCCC TTAATCTATT GTGGTTTATA TCAGAGATCT  
 36951 CTCAGCTGTG TCAGACAGGC CATGACTTAA GTGACACTGC CCTCTTGATT  
 37001 CTCTTCATAC TTTTCCAAC ACAATTCTTT CTCCTGGGGT TGCTCATCTT  
 37051 AACATAGCTG TATCATTTAT TGTAGACACA AGGTCACTTT TGAGAGTGAA  
 37101 TGGGACTATA TTAATAATTG TTCCAGGTAT TAGGTGCAAA CCCTGGGCAA  
 37151 TGCAATTCAT CCTCCATCTC CTCCTTATAT TTATGTGTTT ACCAAGTTGT  
 37201 TTTTCCTGTA GACTTTTTTT TATCCTAAAC CCTTTTCTA TGTCTCATT

26/124

37251 CACAACCTTTA ATTCTAATCT CTCAAATCAA CATTTCACTT TCTGTCTGAG  
 37301 ACCTTTTTTCA GCTCTAAAAC TAAAATCCCA TCAGTGTGCT AGACCATATA  
 37351 GCCACCTGAA ATCAAAGTCT TTTCTTAAGT TCTTTTCTTC TATTTGTCTT  
 37401 ATAATTTTCAT GTATCATCCT TCTCTCTACT CTAGCACAAA ATCTGTGTAA  
 37451 TCAATAGTCT TACTTGAAAC TGTGCTCTTC ATATTGTACA TTTTCAATAG  
 37501 ACAGGAACCT GTGATTTTAT CTTCAGAATA TCTCCTACAT CTGTCTCTCA  
 37551 TTTTCAGGGA CATTTGTCCTT GCTGAAGCTT TTTTAACTAT AGACAATTGC  
 37601 AGCAGATTTT AAACGTATCT TACTCTGTCTG ACTCCCTTAT GTTTCAACAT  
 37651 TTTCACCCAT TGGAAGGTAT AAAAGAAGAT ATTCTCTGTCC GTGTCAACAT  
 37701 AATCTCATGT ACCTCTCCAG ATCTTAGAAA CACGTATGGC TTCAAATCAG  
 37751 GCATTTGGAG ATCTTTATGC TGTATGGTTT CAGAGTGGAA AAAATGATTG  
 37801 ATTCAAAAAC ATAATATTTA AAGAGTTTTT ATTGTATTTA CAGTTCACCT  
 37851 GAACCTCTGT TCATTGGGCA AGAAAATGAG TACTCTTAAA ATGCAATAAT  
 37901 AAATTAAAGT TACTTTATTA TTAAATTTTA AATATATATA TATATACTTA  
 37951 CCTTAAATAT GTCCTCTTGT TGTCTTTTAG CATCACCCAT TTTTGATTG  
 38001 ACCATTATCT TTTCTGAATA ATCAGTAAGA TACAGGATTA TTATTAATGT  
 38051 TCAAAAGTTG CAGTATTCAT GTTTTCTTTA TTCTTTCTAC CAATTAAAAT  
 38101 GTGTTAATAT ATAAAATTTT TAGAAAATTTT ACTATAAAAA ATCACAACAT  
 38151 ATATTAGAAA ATTAAGATCA CTACAATATG TCATATTTAG TAGACTACTG  
 38201 TGAGCTACTG CCACAGTAAA CTATGGTTCG TGTGTCGTTC CCAGCATGCT  
 38251 AGCCCTAGTA GAAACCATTG CCATTCAAGA AAGACTAACA AAGTATAGCT  
 38301 TACATAAATC AAAAAGTCTT TGGATGAAAC TTCATTTGGG AAAATAACCC  
 38351 AATCGCTACC CTTCAATTTT TTATGAATGA AAAAATGGAA GAATAAAGGC  
 38401 CTCTAAGATC CATTCAAAGC CAGGAGACAC ACAAGAATTT CTAAATAGAA  
 38451 GAGAAACAGA AGAGGTCATA GTTCTTGTTA GCCATCTCAT AACCTGGTGA  
 38501 GACTCATTGT CATGCCTCCA TGCATGATAA CAATCGCTCA GATTCATTTT  
 38551 TCATCTTGCC ACAAGGGTTA CATGCAGGAA CATTAATGTC AACCTGTCAC  
 38601 TTCTAATATC CATCTAATAT TCTCTAAATT CGATGGATCC TTTTGCATAT  
 38651 GGTGATTGTT AAACACCTTT GCATAGGAAC AGTTTCTATG CTTTGTACT  
 38701 CAAATCTTCC TCTACCTTGA ATCCTTTCCC ATCTTCGTGT TCAACCTTCA  
 38751 ATCTTCTCAG AATGAACTCC TGTCTTCTAT TCTTTCGGAA GCATAGAATC  
 38801 TCACGGTCAG AAGAGACCAC ATCTGGTTCA ACCCTTCATC TCTTATGTAA  
 38851 AATTTTATGA CATCTCTAGC TTCTTCTTTA AACCACCAA TGACAGAAAC  
 38901 TACTAAAATC TAGAAATAAC ACCTTTGAAA TTCTTTCTTT AAGAGATCAA  
 38951 ATAAAATTTT CCTGAATCTT CACCTATTGT TCCTAGTTAT ATATATCCAG

Fig. 2 (cont'd 21)

27/124

39001 ATTCTACAAA ATAAGTCAAA GTTAGATTGC ATATGACAGC TCTTCATATT  
 39051 TAAAACAATA TAATAAACTC ACTAGTTAAT GTCTAGCTGT AGATGCAAAA  
 39101 GTAGAGAGTG ACTTGGGGTT ATTTAAAAAC CCAGTCCAGC CAGACACATT  
 39151 GGATCATGCC TGTAATACCA GCAGCACTCA GGAGGCTGGG GCAAGAGGAT  
 39201 CCCTTGTCCA GGAGTTACAG GCTACAGTGA GCTATGATCG TGGCACTGCA  
 39251 TACTCCAGCC TGGAAGACAG AGTGAGACCC TGTCTCACAA TAATAGTATT  
 39301 TAATAATATC ATAAAAACCC AGTCCACATT TATATAGGAT CCTGTTTTCC  
 39351 TCAAGTTACT ACAAATAAAT ATATAATCTT AATAAAAGGT TAGTGGCTTT  
 39401 GCCAAGATAG TGGCTTGGCT ATGCAAATGC AATTTAAGAC AAAGTTGGTA  
 39451 GCCCTCTTTT TCCTAATACA TTGCCATATC TGTTTCTCTT CTATTTGGAA  
 39501 ATTCTTGTGT GTCTCTTGGC TTCGAATGGA TCTTATAGTC CTTTTATTCT  
 39551 TCCATTTTTT AGTCATAAAA AACTGAAGG GTAGTGATTG GGTATTTCG  
 39601 CCAAAGCAGA TGGAAGCAA AACTACCACT AGAAGCTCTT TACCAATTTG  
 39651 TGTTCATTTC AAAAAATTAT CTTTGTATGT CTTACATTTG TCTTCTACTG  
 39701 TATAGTTTTT CTTGTTCTAT TTTACATATT AACTTTTCTC CTTCTTCAGA  
 39751 CATCTGCCCT ACTGGCTACT CTTGAAATCA GAGACTGTGT CATATTTTTT  
 39801 CTTCTATTCA ACTACAACAT CTAAAAGCAG ATCTGTCATA GTTATTAACT  
 39851 TAATGAACA CTCTTAAATA GTTAGGTGTA ATTTCCAATG CAGAAGCTAT  
 39901 CAAAAGGGTT TGTAATGCA AACTATTCCC TTTAAATCT ATCCTAATCC  
 39951 TCATTAATGT TTCATCTTGA TAGAGCTAAG TATTATGTAT TGAAATTGTA  
 40001 GAAGTACACT TCACTTGAT ATCTCTGCAA TCATTTAGGT AAGAATTATA  
 40051 CAAAGCCAAA AAGCAAATAA AATATCCTCC TAACCCTATA GATACGTATA  
 40101 CTAAAATGAT GCACTTGCAA ATTTGTTTAA TACTTCATTA ATTTAAACAA  
 40151 GAGTAAATTC ATACTGTGAA CCAAGAATAG GGTGACTTAC CCAATCTTG  
 40201 CCACCTTAAA CATAAACATT TTAAGTCTTC AATGTCCTAC AGTGACCTA  
 40251 CTGGCTGTTG TCACTAATCA GACCGAAATG GTACTAATGG TCACTGCAGG  
 40301 CTGAAGGAAT ATGCTTGAAA GATAGGCAGA TCCTCTCCCT CTCCCTTTTT  
 40351 TACTTTTTTC GCCTTTCCAT CCTTCTTCT TTTTTTCCAA TAGATTGTGC  
 40401 ACTTTGGAGA TTCATATTTT CTTCTTTTC CATTACATTT TAAATATGTG  
 40451 ATTCTTAGTC CTATGCTTCC TTTTACTCCA ATCAATAACT GGCTCTATCA  
 40501 GAGGGTTGTT CTGTGTGTTA ATTCGGTTAA TACCAGGATT ATCAAGCACA  
 40551 GTGCCTTCCA AATGTGAGAT ACTTCTCTCC GGTACCTCT GGGTTTACTT  
 40601 TTCCTGTTTT ACATTGTTTT GAGAGCCAGT ACTTGATTA AGAAGAAGTT  
 40651 TAGTGCCTGT GTCACAGAAA AAATCTTAGT AAATTTTGAA GTGATGTCAG  
 40701 AAACAACCTA AGCCACTGAC GGATTCCACA GGGTTTGA AATACTCGTT  
 40751 AGTTCCCTTT ATATCTTAAG AGGCTCCTGC CTGCTTTCTC ATATACCAGT

Fig. 2 (cont'd 22)

28/124

40801 AACAACTTG CTTTTCTTAA ATATGAGCAT TTAGAATATC TTTCTCAATT  
40851 TTTCTGTTTT GCTTTTATTC CAAATTTTAC AACTATATTG TTTTCCAATG  
40901 TAGTTGTACA TACAATCAAC CAAATCTTTC CTAAATTTGA TGACTACCAG  
40951 GTGAGGACTC TTTGGCAATA AGCAATAAGA AAATAAATTG TTATTAAAAA  
41001 TTACAGACTT AAGATACTTC TTTGGAAATA TAACATGTTT GTGACTTTTG  
41051 ACCATCTCAT CATGATATGC TCATCTTAAA CAGAGTAGAA AATCATTTC  
41101 TATAATTAAC TTTATGGTGG GCTGCAGATA CCATGTATGT TACATTGTGT  
41151 TTAGTTATAA AAATGTTTAT TATACACTAT TTCCTTATAA TCTAACTTTG  
41201 ATAATAATGA TGGTCCTAAT CATGAACCTA CATCAATTAA GAGCTTGAAG  
41251 TGACTGAGAG TATTTGCCTG GAAGCATTTA AAGCCCTTCT TGGGAAATTT  
41301 AGATGTTTTA TATTTTACTT TCTTTTGTAT TTTGCTTTTT CCATTAAAGT  
41351 GATTACTATT TTTAAAGAGA AAACCGAAAA CTCTAGAAAG ACCATCTTTT  
41401 CTTCATAACA GGTAGCAGAA AACACCATGT TATTACATTT CTAGCAAGAG  
41451 CAGTAGAGGT GACTTGTTGG TTTTGTGTAC TGTGCTTTA GAAATTGATG  
41501 TAAGGCTTCC CATAAACGTG CCAGAGGAAA AGAGGGACGC AATGGGATCT  
41551 GTTATTGAAC ATTTTCAGAG CAGACTCTTA CCTTAAATAG GGACTCACTA  
41601 TACATTCATG TTTTCATAAG TATTGGGATC ATGTTCTTAC TTTCTATCAA  
41651 CCTGCTATTT TCATCTTTCA AGCTTAAGAG TAATAGGCTC TGTGTGTTTT  
41701 GTTTTTCAGT GAGCCCAACA AATTTGTCTC AATTTAACCT TCCCGGGCCC  
41751 AGCATGATGC GCTCAAACAG CATCCCAGCC CAAGACTCTT CCTTCGATCT  
41801 CTATGATGAC TCCCAGCTTT GTGGGAGTGC CACTTCTCTG GAGGAAAGAC  
41851 CTCGTGCCAT CAGTCATTCG GGCTCATTCA GAGACAGCAT GGAAGAAGGT  
41901 AAGCGTTGAG GGGGATTAAA GATGAAGTCA CTTTATTTAA ACCCTGAGAG  
41951 GGAAACCATC GTGTCACTCA CATCACAAAG ATTCCTGAAG AGGAAAATAA  
42001 ACTAGTGTA TATCATTTG GGAAACTAGA AGCTTGAAGA AGTTTTATTCT  
42051 TGTATTATCT TCTATTTCTT TATGTATTTG GAAATATGCC AGAATTTGTT  
42101 TATATTAATA CTTGGCTGTA GAAGAGTTTA GACTAAATCT ACTTTTCCAA  
42151 TACAGAAATA TACATATAAA CTATTTTCCC AGGTGCATCA AATATCAGAG  
42201 CAAATGTTTT GTTTGACATT TTGGTTAAAG AGCCATAAAG ACACACAAAC  
42251 CAGAAACATT ATTTTATGAA AATACCACAT GTTGCTGACT TTTATTCCCA  
42301 GGAATTCCCT CTGGTGCTAA TTTTTTATTA TATCATTTTA GAATTCATAT  
42351 TGTACCTACT TTTTGTCTT ATAAGTCACT ATTTCTTCAT CCAATGGCAA  
42401 TAAATTTGTC ACCTAACCTA ATAAATATCT TTATAGTTAT ATAGTTCTAT  
42451 GTAAATACTC CAAATAAATC AGCTTGAAAA CCTCAGGAAG CTGAGTTGAT  
42501 GCTCAAATAT ATATATTTTT GTAAACTGTA GAAGCTCAA TGTCAAATTT

Fig. 2 (cont'd 23)

29/124

42551 AACATAATT TGAGAGACTT TTCTCTTTGA TTTAATGAAT TTTTTTAGTA  
 42601 TCCATAAAGA AAACCTACAG CATACATATT ATAAAGCATG TCAGCTAAGG  
 42651 ATAAAATAAA ACTAGACATA CAAATTCAAA CTGATTAGAA TGAAATTATT  
 42701 AACCTAATA ATTATGTTTA AAAGAAAAGT CTCCAAATCT TGAGACATAC  
 42751 CAGAGTTTAA GTCTTCAGCC ATCCATTAC TTGTGGTATA AACTTAGGCA  
 42801 AGTTTCTTAA CCTTCTTATC CCTAAGTTCT GCATCTGTAA CTTCCTAGGT  
 42851 TTGTCACAAG GATGAAATAT GAGAACAAAG AATAATTCTG TTCCATGATC  
 42901 TTTTCCCTTC CTACCTTCTT ATTTAAAGTA TCTTCTGACT GAGGGGTTAG  
 42951 GCAGCAATGA AAATTGACTC ATGTTTTTCA GGTCACCACT ATGGATTCAA  
 43001 TATACTGGCA TTAAATCAGT AGAGAATAGT TGTCATTGCC TTTTGCAATA  
 43051 TTAACCAAAC CACTCAGTTC ACTGTGACAG ACAGTGAATT ATATCCAATG  
 43101 ACTCCACTGA TTTTTTCCAT GTAGATAGAC AAAATATAAC TACTCTCAAA  
 43151 TGTAAGGACC CTGCTTTCTG AAATGGTTCT GTTGCTCTCT TCACAGATAG  
 43201 GCTTCTTATA ATACTTTTAA AATAATTGTC TAAGCATACA GATGGCTTTC  
 43251 TAGAGTGTGG CATTGACAAA TAAAGTGATT TTTATATACT GGGAAATTCT  
 43301 GGCTTCAAT GTATCAGGAT TAAATAATCT GAATTTCTGA AAGCTAGCCT  
 43351 AAGTGGGCAA GATGGCTTTT TTGTGCTCAC GCATTGAATA CTGAACTATT  
 43401 CTAGTTCTTA AATGGCGATC TAGATTCAAG ACTTATTGAA CTAGATTGAA  
 43451 GGGACTTTAT TGATATCCTA CCTAATGCTC AACTGACAG ATGAAGAGAC  
 43501 TGAGCCACAT GTTCTAAGGT CATAAACAGA AAGAATGAGA ATGAGATGGT  
 43551 CTAATTAATT GTCCACCTTT CCTATGGTAC ATCAGGGTAA CACTTTAGTT  
 43601 TACGAGGGTA TTATTAGAGA TAGAAAGAAT TTTTTTTTAA ATAATTGACT  
 43651 CAAATACCAA CATTTTGCAC ATTACATAGA GTAATAGCTT TGCCCAAGTT  
 43701 AGAAACTGG GGGTCTTCT TTATTCCTCT TTGACCACA TCTATATACT  
 43751 CAGTTTTAAA AAGGTCTTCT CTGGTATCCT TCAATTCCAT CCCCATGTTT  
 43801 TCATCTACAA GCCTAGTGCA GCTATTCCAG CCGTCTCCTG ATCAGGTCTT  
 43851 AAGCACCTCC CATATGTCCT TGTAGTACCC ACCATATTGA TCTCAGTAGC  
 43901 AATCACAGTA CTCTATTGTA AATATCTTTT AAATTATTAT CTTCCTTTG  
 43951 AGCTTTTGGG ATTTTATCTT ATTTATTTTT GTAGTTCCAG GATCTAGCAA  
 44001 CAGCTTGTC AATCGTTCAT ACTCAACTAA TGTTTGTTTA ATGCACAATG  
 44051 AGCAGAAATA AACATACTAC TCCATAGTAA AAAGAGGATG AACTTTTCTG  
 44101 CAAATATTAA TCAGCACCAT TTTATCCACC TTTGGGTTT AGTACATTGG  
 44151 AAGTATAGGA GTATAAGCA GAATGTCCAA TGTTTACAGT GATATTTTGA  
 44201 AATAGATAAA AGCCAGTGCG ACATTTCCAT TCTCAATTTT TCTGAGACAT  
 44251 CACTTGA AAAAAGTA TTTTCTCTT CCTAAAATTA GTAAAGGAAC  
 44301 AGTAATTCCA CATTTATAAG AGTATGATCA ACGCATCACA GATAATGTTG

Fig. 2 (cont'd 24)

30/ 124

44351 TAATAACACA TTAGATAAAA GTGCTTATTT TCCTGAAATT ATATGGAGAA  
 44401 AAAAATCTGA AAGTGGACCT TTGTTGGATA CAAATGAAAT AAATAAGGTA  
 44451 CATACATTTT TTAAGGTTTC AAAGTTTATG GCAACTTTAG TTTGGGTTTC  
 44501 CATGCTATTC TATTTATTAT ATGGGAATTT ACTGTAGCTT TCAACATGTA  
 44551 CGAAACAGGC TGGTAGGGCT CATGCTTGTA GGCTTCTGTC TAATAACTTG  
 44601 GCAACTGAGG TACTTTAGGG AGTATGGATG GGGCTCTTCC ATGTCTCAAC  
 44651 GTCCTGACTG CCAAAAAATT ATAGCAGGCT GGTTCCTCAGA ATCTTATAGT  
 44701 TAGTTGTTAT TACTTAATTT CCCTAACCAC CCGTCTTTA CTTTTTCTGT  
 44751 AAAGGCTGGA ATTTTTGAGT AGACCTTATT GTTTTAACTC TATTGTTCTG  
 44801 TTTGTTTCTT CCAGTTCATG GCTCTTCATT ATCACTGGTG TCCAGCACTT  
 44851 CTTCCTTTTA CTCTACAGTA AGTAATGGCT GTTAAGAAAA AGCTTGTGCT  
 44901 TTTGCCATGC ACACAGATGA TGAAATAGAT CATTTTACTG TGAACAGATC  
 44951 ACATTCATCT ATGACTTGCA CAGGAGTTGT GTAGCAAAAT AACGGCATAAC  
 45001 TCTAAGCTGC CCAATACCCA ATAAAGTGCC AGGTGCTCCA CCTGCCATTC  
 45051 TTTGGTCACT TACATGTGCT TTCACTTGGC TTTTGTGCAC TCATCATAAT  
 45101 CAATGAGTGG ATGTAGAATT CGATTTTATA AAACCTACTG AGGTATGACT  
 45151 TGGAGTCTCT GAAACCATGT ATGTAGTCTG CTATACTATC ATTTTAGTAA  
 45201 TGACGAGTTG TCCATGTTTT GTTCTTTGAG CCGTGACTGT TAATTGTTCT  
 45251 ATAGTATTTT CTTCTCATTT TTTATTTTTA AGTTTATTGT TGAGAGGATT  
 45301 ATCGAAGGGT AAAAGCAGTA AGGGTAAAGG GTAAAAGCAT AAAAGAACCA  
 45351 GAGATGTTTT TTTTTAAATA TACCTTTTGA AAGAGTGTA TTTTTTTAAC  
 45401 TTTTATTTTT ATTTTATTTT ATTTATTTAT TTATTTATTT TTGAGTCGAG  
 45451 GTCTTGCTTT GTCACCCAGG CTGGAGTACA ATGACACAAT CATAGCTCAC  
 45501 TGCAACCTTG AACTCCTGGG CTCAAGTTAT CCTTCTGCCT CAGCCTGTCA  
 45551 AGCAGCTAGG ACTACAGGCA CGCACCACCA TGCCCAGCTA ATTTTTAAAT  
 45601 TGTTTTAGAG ACAAGGTCAT TGCTATATTG ACCAGACTGA TCAATACCCA  
 45651 TGGCTTCAAG CAATTCCTCC TGCTTTAGCC TCCCCAAGTG CTGGGATTAC  
 45701 AGGTGTAAGC CAGCACACTT AGATAGAAAC TTTATTTATT AAGAGAAAAA  
 45751 TACCAGTGTT TCAAGTTCTT TTGCAAACGT GTGACATTAT AATTCATTTT  
 45801 TGACAAGGAG AGTTTTTCTG TTTGGTAAAT ACAATTCTAT CTTTTTTAAA  
 45851 AAAGTAGCCT ACAGGAAGTT ATATTTTATG AGTGAGTCTT TTTAGAGCTA  
 45901 GGTTAACAGT GAGGTATATT TAAAAGCAGC CTAAGTGAATC TCAATGGGAC  
 45951 TTGAGTACTA TGAATAAGCC TTAATCCTGT ACTGTAAGGT TCATGAAGAG  
 46001 TTCATAGCCT CTGCTGTCAC TGATCAACTG AGCATCATGG GCAGTATTTT  
 46051 TTTCACATCAT TATCATTAGG TTCAAATGTT TGTTTGAACC TTCTCTTTAT



46101	AGATTAATCT	CATATATT	CTGCCTTACA	TAGTCATTCA	AAATCTGACT
46151	GTTATTGGCA	GAAGTAATAT	TTTTCTAATC	TCTCCTTTCA	ATGATTAAAA
46201	TTACCCATAG	CTTCTAGAAA	TTAAGAAATC	ACGATTAGTT	TTTAGGTAAA
46251	TGTACTTTTT	GTGCAAATGG	ATAAAGTGAG	GAATGTGTAA	ACACACATGA
46301	AAAAAACACA	TAAAAGAAAT	ATATTAAAGAC	TTAGTGTTC	TCCTGTTGGG
46351	CCAGCACTGC	CATTTGTTGG	GGAATTGTAT	TCTGATTTAA	ACCATTGCCA
46401	TTTACATCTA	TGTGTAACAT	CAAAAGATGT	AGCATCATTA	TTATTCTAAA
46451	TACATACAAT	AATTAATATT	TGGATAAAGC	TACCTTCATG	AAACCTAAGA
46501	AAAACATAAT	TAAAAAGAAA	GAAAGAAAGA	AAAATACACT	TAGATAGAAG
46551	AAATAAGGTC	TAGTGATTGG	TAGCACAAATA	GAGTGACTAT	AGTTAACAAT
46601	AATTTATTGT	ACATTTCAAA	ATAGCTAGAA	AAGAAGATTT	GGAATGTTCC
46651	TAACAGGAAG	AAATGATATT	CTTCCTAAAT	GAAGAATGGG	ATATTCCACT
46701	TTCCCAGATT	TGATCGTTAC	ACAGCATATG	TTTGTATAAT	ACCACATGCA
46751	CCCCATAAAT	ACATACAACCT	ATTGTGTATC	CCAATATTAA	AGATTTTTTT
46801	GAAAAATTTA	TTCTCAAGA	AAAGGATCAT	GAGTTTAAGA	AAAAACAGAT
46851	TACTAGTCTA	CCAGTGTCCA	GTAGACCTTT	CTGTGTTAAT	AAAAGTGTTT
46901	TGTATCTACA	CTATCTAATA	TAGTAACAT	GAACCATATG	TTGCCATTGA
46951	TTATTTGAAG	TATATCTGGC	AAAGAGATGA	ATTGACTTTT	TTATTTTAAT
47001	TAATTTACAT	TGAAATAGCC	ACATGTGCCT	AGCAGCTACT	AGATTGGATA
47051	GTGCAAGTTT	ATAGAGAACA	CAAGGGGTAC	ATTTGTAGAT	AGGAGTGGGA
47101	TGTCAAAATG	ATGAGGATAA	TTAGAAAGCA	TACATGAGAA	ATATTGTTTT
47151	AAGAGTAGAA	TATGAAATGG	GAACACAGAT	TAAATAGAG	TATGTATATA
47201	TATACATATA	TATGTGTATA	TATATACATA	TGTATGTGTA	TATATATACA
47251	TATATATGTG	TGTGTGTATA	TATATATATT	TATAGGCCAA	TATATGGAGG
47301	TAGGGTATAT	CCTAGTGTTA	AGTGAGTAAA	GAATGGATTA	GGTGATCGAG
47351	CCACATGAGA	AGGTGATATT	ATTAGAAAAT	TGAAAGTTGT	ATTTGAGATG
47401	ATGAAAATGA	TATATTTGAA	TTGAAAAGTA	AACTGTAGTA	AAATAATTCA
47451	AATAAATGAA	TATTTGGGGA	ACTACTTAAG	AGAAAAATCA	TAAAACATGA
47501	GGAGTCATTC	TTTCCCCAGT	CCGCCATGAT	CAGGCCTTAG	GATTTAATTG
47551	GCAATGAGAA	AATACCTATG	AAAATGCTTT	TTAAACTATC	ACATGAAAAA
47601	GCAATTTATT	ATTTTTCATG	CCTTCTTAAT	AACTCTCAAT	AGAGATTTAG
47651	TTGATTTGCA	TTTTTGCCTG	GTTCAATCAA	GAAATTATCG	CGTGACATCA
47701	GGCAAGTTGC	CAAATTTCTT	TGGACTATAC	CTATAAAATA	AAATTTGAAA
47751	ATATTAGCTA	GATCTAACCC	ATTTGTCTCC	GGATGTCTGC	AAAGTGGTTG
47801	GAAATCACAA	GCCTAACCTG	ATCTGCAGAG	GTGTTACCTT	TGGCAAACCT
47851	ATGGTTTTTG	TGTTTGTGTT	GAAATCTAAG	GCCAAGCGCG	GTGGCTCATG

32/124

47901 CCGGTAATCT CAACACTTTG GGAGGCTGAG GCGGGTGGAT CACTTGAGGT  
 47951 CAGGAGTTTC AGACCAGCCT GGCCAACATG GCAAAACCCC GTCTCTACTA  
 48001 AAAATACAGA AATTAGCCCG GTGTAGTGGC ATACGTCTGT AATCCCAGCT  
 48051 ATTTGGGAGG CTGAGGCAGG AGAATCGCCT GAACCTGGGA GGCTGAGGCT  
 48101 GCTGCAGTGA GCGCCACTGC ACTCCAGCCT GGGCGACAAA GCCAAACACT  
 48151 GTCTCAGAAA AAAAAAAAAA AAAAGGAAAA GAGGGAGAGG GGAGGGAGAG  
 48201 GGAGAGGGAA TCTAAGCCAA CACTGTGAAA TATTGTGAAA TATGGAGCTT  
 48251 CTACCTAAAA ATTCAAAATT TTAAATTCCT TTTAAAAATA ATTGGAATAT  
 48301 CTATGGAATA TCTAGCAATA CTAAGATGAA ATTCTCTGG GTTTTCAGTC  
 48351 ACCTGTAATT GACACCTTTA GATGTTGGCA TGGGCTCTCA GGAAGCCACA  
 48401 GCCTCCACCA ATGCTTTTCT TCCTGACACT GAAGCTAAAT TTGGGTGGCT  
 48451 AGTTTTCAAT GTGCTGTTGC TTTCTCATG GGAAAGAAAT ACCCTTTGCT  
 48501 ATTTATATTG CTGTCAAATG GAAAAATGAA AGACAGCCAA GGAAGATCAT  
 48551 GTGACTATTT AAATACTTCA AGTCCATTTA TTCCTTATTA GCCTTGTCCT  
 48601 GTTAGGCATT TAAATTTTTG ATCCCTGCAA TAGATGTTTT TTGATTAACCT  
 48651 GTATATTAAA AACTATATTT AACCTGTTTT GAATTTGAAT TCTAAATTGT  
 48701 ATTTTTTCAT GAGAGCAAGT GTCATTTTTG ATTCATTGTG GATTGTTTAA  
 48751 CATGTTGCCT AACAAATAGC TAATACTAAC GTCATAACTT TTAAATTAGT  
 48801 AAATTTGAAT GGATAAATGG CCACTTATTG GCTTATAGAA TAAATAAAAA  
 48851 CATTTTTATT CAGTCAAGTG TTTTCATTTT TTTATCATCT CCAGGACATT  
 48901 GGGCTTGCTC AAAACCATTG TAAAAAATAA AATGGCAAAT AATCCAGTTC  
 48951 CATCATGATA TCATTAATCC CACACCTAAG CTACTGAAAA AAATATATTA  
 49001 ATATTCTGGC TCATTGCTTT ATTTTTATGG TAACACCCAC CTGGTATTAA  
 49051 TAACCACAGA GTACGAAAGA AGGCAAAGGT TAAAGCAAAT AATAGTTTTG  
 49101 AAAAATTGGT AGTGAAAAAA GTCATGCTAT ACGGTATGTA TATAATAGAT  
 49151 ATTTAATGAT TATGCTTGCT ACTAGTATAT GTAACAGGAC TATTATAGAT  
 49201 TAACAAAAAT GCGGTGAGTA TATTTCTTGA TTATTTTTTA AAAGAATAAA  
 49251 TTATTATTTA AAAATACATG AATTATTTAT TGATTCTTGA ATCTTTACCA  
 49301 GCTTTCTATA ATTCTAGGAA GCCTAGAAGC AGAATTGGGC AGGATAAACT  
 49351 GGCAAAAAAT GTAAAAAGTA GGCCGGGCAC GGTGGGCTAC AGTGAGTCGT  
 49401 GAATGCGCAG TGCACCTGAG TGATAGATCA AGATCCTGTC TCAAAAAAAA  
 49451 AAAAAAAAAA AAAAGAAAGA AAGAAAGAAA AACAACAACA AAAACAAAAG  
 49501 CAAAGTACTA GGGAAACTA ATAGACATAG TTACATAGTT AATTGTGCCA  
 49551 TATGTTTTAA GGCAATGAAA CTTTTATCTT AATATTCCTT GCTTACTTTT  
 49601 TATTCAAAAA CCAAACTGTG TATAAACCT TAAAATTATT AGGATCTAAA

Fig. 2 (cont'd 27)

33/124

49651 AAATAAAATC TTTCCTTAAA AATCTAAAAT TGAGATGTAA ATTATTCAAG  
 49701 AGTGCTTTTT AAAACAGTTT TCTTATAAAG GCTATTAGGA TTCTACCACT  
 49751 TAGCCACTTT ATTATTTAGC CACTATATTA CTAAGTTTAC ATATTTTAA  
 49801 AGGTAGTGAA AATATAGGGA AGACAAAGCT CAGGTTAAAA GAGTTTCTGG  
 49851 CAAATAAAAT ATATCCTGAT GGTTAGACTA CTTTGCTTTA TGTTTCTGA  
 49901 AAGAAAAGCA GTAAAAACA GTTCAGGTAG TTTTGTGTCA ATTAATCTAG  
 49951 AACTATACCA AAAGTAGACA TAGAAAACGA GAGATTGTTT TTCAGCTTTG  
 50001 GATCTGCTTA TGGCAATAAG CAGACTTGTA CTATTCAACA ACATTATGCA  
 50051 TTCTTCAACT TTCCCAGAA TAAGGGAGCT TCCCAAATGC AATGGTGCAC  
 50101 ATAACCTATT TTCTGGCATT TTGCAGCCCA GCATGAAGAA GAAAAACAGA  
 50151 GCTAGGAGTT TTCTGGAAGT CAAGTCAAAA ACACCCTGCA AATTCCTATG  
 50201 GCAGTCCTCC TTCCATAAG CTGCATAGCC AAAAATGTTT GCCAGACACT  
 50251 TTTATCACTG GGTGTTTCAG TGTTTTTCATT GTTTAAGCGT TTTGCTGACT  
 50301 TGTGATAATT AAAATTATTA ATAATCATT AAGAAAGAAA AAGTAGAAGT  
 50351 AAATAATGTT AATTATCTGT GGTTATCAGT AGAGGTCTGT ATGTTACCCC  
 50401 AGCTTTATTT GACATTGTTT GTGATCAGTA AATCACAGAA TAAAATCTGT  
 50451 ACATCTAAAC CTTGGCTAGA GGTCTCTATA ATTTTATGGA GTCTGTTTCC  
 50501 TACAATCTGT ATGAAAGATA CTTCAATATT TTAAGTTTAC ATGCACCCAT  
 50551 CTTTTTTAGA GTATAATTTT ATAACATTTT GGTTTATGTT GCTTATGATT  
 50601 TACATCTTAG AGTCTTTTAA TTCTGTCTTT TGCTTAAAGG AATATTATGG  
 50651 ATCAAATGAC CTATATTTTA AGAATACCTT ATGGTTTATA TATTAAGAAA  
 50701 CATTTATATA AAATCTAAA GTAACCTGCT TGTACTATTT CAATTGAATA  
 50751 ACTTAATGTA TTTCATTCTA TTCTTCTCAT AGTAGATAAT AAAAAGTACA  
 50801 TCATGATTAT TGTATTCAAT TATACTTGTG GAATTAATTG AAAATAGTTT  
 50851 TTATAGTTAA AGTCTTTCTT TTTATTGTTT TACAGGCTGA AGAAAAGGCT  
 50901 CATTCAGAGG TAAAAAATA TATGCAATAT TTAATATTT TCTATTTTAG  
 50951 TTTGCATTCA TGATGAAATT AGTCTTGTGA CCACTAGAGG GCTCTGTGAT  
 51001 ACAATAGCAG AACTCCACAG GACTGCTGAA GTAAGGCAGC TAATTGATAA  
 51051 ATGGTCTTTG ATATTGCCTC TTAATAATAA AATGAAAGGA AGTTTGTATA  
 51101 GCAAGCTGTC CTTTCACATT CTAGATTGAG TCTTAGCTCA ACACCTAATA  
 51151 AGTTTTCTAT AATAGTAAGC ACTCATTAAG TCATTGATAA ATGAAGGTCT  
 51201 ATGGTCTTCC TATTTTATTA CAGTCTTTTT CCCACTCCCT GTAAGACCAT  
 51251 CTACACAGGA TAATGGTTGA AACTTGGGCA CCAAGCCTCC ACAACACAGG  
 51301 ATACTAGCAT CTCAGACTAT CTGTTTTGTG TCATTATCTT GTTGCCCTCA  
 51351 ACTGCCATTT TATGTGTGGT GTGTACCTA TTGTTCTAAT CACATATTTT  
 51401 ACAAATACAT ATTTGGTTGC ACTCGTGAGC AAATCAAAC GCATTACAGGA

Fig. 2 (cont'd 28)

34/124

51451 AAGAATACTA TTTTAATTTTC CCTTGGTAAA ACATTTGTCC TGGTCAAAGA  
51501 GAGCAGGAGG ACTTTAATTA TGACTTTATT CAAGGTGAGG TAATGGCTGT  
51551 TTGATTGGTT TACACTGAGG CAATCAGACA ACAGAGAAAA AAAATGCCTT  
51601 AACACAGCT TTTGCAAAAG TATTCCTTTC CTTTGAAGTC TTATTTTATT  
51651 AGCCTTTAAA AATAAAATTT GTGCTATGTT TAAAAATATT TGAAAATTAT  
51701 TGATTAAACC AATTGTCTT TATAATCTCT GAACCAAAGA GTGGATATGA  
51751 TTTTTAAAA TCAAAGTGGT TTTATTTACA TCACATGGAC ATGACAAAGC  
51801 TTCTAACACT GATCATAGTA TAGCTACTGA AGCATCGAAA TGCTACATCT  
51851 ATTTGCCTTA GTAGTAGTTA TTCAACTCCC CTTTATCAT TGATGCTGTA  
51901 TCATGAGTTA TGGTTTAAAA AAACAATTTT AATCACTTTA CAGTTTCCTG  
51951 GATTATATTT TAAAGATACT GGAATCATGT AATAGAGACT ATTTAATTTG  
52001 AGAAATGCTC TTTGAGTTTG GATTCATTTA TGAATAAAAT AGACGCTGTA  
52051 TTTTCTGAAA TCATTCATAG TCATTATCTT ATAAATGTAA AGCAAATGTT  
52101 ATTTTAGACT GGGGTGTATC TGTTCGGAA AAAAAAAAAA ACAGGAACGA  
52151 AGTAGAATCA CATTGCTGA AATTATATAA GTGTCTACTG TTTCCAGCTT  
52201 AGAGTTCTCT ACTTTGTTAG AGTGTTTGAG TTGACCACCA TTTATTTTCA  
52251 ACAAATCTA ATGCCCCGGG CAAAACTAG ACAGTTAATA AACTATGTCA  
52301 AGAATTCTCT TTCAAAGTGA GACAGCATTC CAAAAGTTCA ACTACAATA  
52351 TAGATAAGAT TTGTTTTTGA AGAAATGAGA AGCATCAAAA GTAGAATGTT  
52401 TAACATCCAA GTAAGTAAA TCCCTTGAGA CTAGATATAT ACTTATAGAA  
52451 CCTAGTGTCA GATTGTTATA AATGTTCTAT CCTTATTAGT CACAACATGA  
52501 GACTTGCAGA ACAAACTGCA GAAAGTGCTT GAATTAAAC TTTAAACATG  
52551 ATATAATATA TCCTTACCCT TTTCTGTTTC AGTTTATTG GAGTGTGAAC  
52601 TTAAGTAAA AGAAAGATAC CTTAGAATAT ACATTATATT GGTTTATCTA  
52651 ATTAGTTGCA CCTATCATTG GTTTTTTCCC CTGATTTTTA AGATGTGGAT  
52701 AAGCTATAAA GCATCTCTGA GCTAATAATA ACTCACTAAA TAAAGGTCTT  
52751 GATAATACAG ATTTGGGAAG GCTTCTCTGC AGTCATTGAA ACTCCAGCCA  
52801 ATAACAATTT AAATGTGAAC TGATTAAATG TTGAATTAAG CCAAGTTT  
52851 AGTGATTGCA GGATATTCCA TAGCCTTTGA GAAGTTTCA AACTATGAGA  
52901 AATTAAATG TACAGAGGAA AAAAAACCT AAGATTTTCT GAAAAAGAAC  
52951 ATGGAGTATC TTTTACTAAA AAAGAACAAG AAAAATATGT GTGTATATAC  
53001 AGTTTTTATA AAGAAAATAT TTTTCTACAG TTTTATTACC ACAGTTTTTC  
53051 TAGAAGGAGA AGAATCAATA CAGAGGGTAA ACTGCTCTTG AGTCATTGCG  
53101 CATTTGAGGG ATGGCAAATG GAGCAAGTGA GCGTACTTTG ATTTGTAGAT  
53151 TAGAGTTTGA CACATAACAC TTTGCTTTTG AATGACATTT GCTTGTTACT

Fig. 2 (cont'd 29)

35/124

53201 GTGGAGTCAG TGTTCATATC CTTTATTTTC AGGAGTTGCT GCTGATACAA  
53251 TGGGGTTAGA ATGAGCTAAA TACAGCATTT GCTTCTTGG TTTGAATTCT  
53301 GGGTTTTAAG TAAAAATCTA CTTGCCTATT CCATTGATTT TTTTAATTGC  
53351 ATTCAGCAAA TCCATAAACT GCGGAGAGAG CTGGTTGCAT CACAAGAAAA  
53401 AGTTGCTACC CTCACATCTC AGCTTTCAGC AAATGTAAGT CACTTCATTT  
53451 TTAAAAATATA TTACAACAAA TTTTATAGA GGAAATGAA ATCATTTTAG  
53501 TAACAAACTT ACAAATTTTC AGTGCCTGAT ACAGACTTAG ATTACCAACT  
53551 AGCAGGACTC ATAAAAAGTT AACATTTTTT GCCTACTCAG TAATAAAATG  
53601 TAAATCCAAA CTGATGAGAG GCAGCAATAT GGTAAAAATG GCTTGTGTGTT  
53651 TCTAATAAGA TTGGAAACAA TAGTAACAGC CATATGGGTT ACTTCTTTTC  
53701 TTGTTTGCTA TTTTATTAC TCCTCTTGCA TAAGATTCCC TGACAATGTA  
53751 AGAGGGGTTG TTAGTGTTTG ACTTTGGAAG ATAAAAATATT CCTGTGCCCCA  
53801 GCCTCCTTCA TCTCAATGTA TTGAACAATT TGTTAAGCAT CCAGTTAATT  
53851 CTAAATATG AAATTAGGTC TAAATAGGGA TAGCTTAGCT GCACTGTGGA  
53901 TGAGATATGG TTTGCTCAAA AAACCTTGGC AGCCTTCTCA TAGCAATTTA  
53951 AAAGGGTACA CTTTACTGG CACCAGAGCA GCCCAGGATG GCAGAAATGA  
54001 TGACAATGAA GACCGTCAAT TAAATTAACA TTTACTGAAT ATCTTCCACT  
54051 GTGTCAGGGA GCACTCAGAG TAGATGCAGA ATGATAAAGG AGAAATGTGG  
54101 CACTGTTCCC AGTCCTGAGG AGCAATGGTG TTAAGAACAG CAGTGAGGGG  
54151 TAAGGAAATG CCTGCTATTT TGCCATATGT CTTACCTCTC TCACTCAACA  
54201 GTCCTTTGCT CAGTCTGCT GCATAGCTTT GGGCCTGCTC TGTGCCTCCC  
54251 CACCCCTCCC ACTGCTCCTC TACTGAGTTT TTCTATCTCC TAGACAAAGC  
54301 ATGATATGTC AAGAGTGAGC AGGTGCAGAC CCACAGTGTA AGACTTGAAT  
54351 AAGAGCCATT TTTAAATTTT TTTTAAGCTA TCATTGTGCA ATATAAATTC  
54401 TAAGTATGTG TATCATTTCA TTCACAATGT ATTCATTTTA GCACTGTATT  
54451 TGAATTGATT TTATTTTCTG AAATTTGGGA GAATTAATTT TGGATTTATT  
54501 CTATTTATTT TTAATAGATG GTGTTAGGAG ATTCCTGAAA ATAATAGCAG  
54551 TTTTLAGATA ATTGTTTAAAG CAATATGAGA AAATAAGGGT ATTATTTAAC  
54601 CTTGTTGTGT TTTTAAAGAG ATAGTCCAGA GGCAACCGTA AATTTTATAA  
54651 TATAGGCTAC ATGTATAGAA GTATGAAATA TTGTTGTCTA GGTTCCTGAA  
54701 TTTGTACCCA GAGGAAGTAG AATAATGTAA ATGTCAGAAC CTCCTGGGTT  
54751 GTGTTTATCT GCAATAAGAA AGGCTCAATG GCAAACCTTA TTTATTAGAT  
54801 TGTCAAGATA CTTGCAGATG TCTTGAATGA TTACTCAGGG TTTCATTTTA  
54851 TTTTAAATGT CCCTTGGTTG AGCTCATCAT ATAATTCAGA TATTGGAATA  
54901 ATAAATGGCT GCTAGACATA GTGGAAGATG GGCTGATACT TTCCATTTGA  
54951 AATGTAATGA TGCTTATTGT CTTCAAAGA AAAAATAAA ATGGTATTTT

Fig. 2 (cont'd 30)

36/ 124

55001 ACATTTTTTT GTTTTGTGTT TTGTTTTTTT TTCTCTGAGA ATCTCATTTCT  
55051 TACTCATGAT TATTGGTTTC TTGTGTACCA TTTCAACATT TTTCTATTAT  
55101 ATGCTAATGT GTATATATAC TTAATACACA CGTGCAAAAG CTTCACACA  
55151 CACACACACA CACACACACA CACACACACA CACACATACA CACACATACG  
55201 GAACCAAATT CTAACATAGG GGAATAATCT TCGGAGTGAA CTCTGTGCTG  
55251 CTGTTTGAAA ATGGAGATAT AATTTTAGAA AGGTTCTGTC AGTTGGCTAC  
55301 CCACCTCGTC TGCTCTAATT ATGCTTGTC CACTATTTTC ACTGATGTGT  
55351 TTTCATGACT TTAGGGCATG AATTCCTCAGC TGGGTGTTAA TATGACCAAC  
55401 AAAGGGTGAA AACAGGTTCT TGCATTTTTT TAAGTACTCT TTTTATGTGA  
55451 AAAGCACAGA TATGCAGATA ATACATAACT GAACATCCAG CATATCTGTG  
55501 GCTTTAAAT ATCACGAAGA AGAGCACAAT TAGGGAAAAG AAAACATCTA  
55551 TAGTGTTCCT CTAGGGGAAC AATCATTTAA AAAAAATAA AAATAAGGAA  
55601 CACAGACTAG AAGCAGCAGT GCCAAATAGA TAATTCATGC TAGTCTTTGT  
55651 GTTAATTTAA AAAGTGCTAG TCTTGAGAC AAACGCCCAA ATTGCTCTAG  
55701 GTTCCACTCA GCTGTATGTG TTATCATTAG TATTAACCTT TGCACGCTGA  
55751 TGGGAGACTG ATATATATCC TGTTTTATGT TCCTTTAAAC AATTTATAAT  
55801 GTAATTTAGA AACCTTCTCA AATCACATTA GATCCACACA AAAACCTGTA  
55851 CATAGCAGCT TTATTTTTTA ATAGCCAAAG AAAGGAAACA ACCAAAAATA  
55901 TCCCTTAATA GGCCAGTTAA TAAACAAATT CTGATACATC TATATCATGG  
55951 ACTACTACTC AGCAATATAA AGAAATGACT ATTGATACGT GCATCAACTT  
56001 GGGTGATCC CAGGGGTATT ATGCTGAGTG AAAAAAGACA GTTATAGAAG  
56051 GTCAAATTTT GTATAATTCC ATTTATATAA CATTCAGAA ATGGCAAAAT  
56101 TAAAGAAACA GAGAACAGAT TAGTGATTGC TAAGGGCTAA GGATGAAGGA  
56151 GAGAGAGAGG TAGTGTGACT ATAGGAAGAG GGAGATCTTT AGTTTGTAT  
56201 TTTGAATGAG ATGGCCATCA CATGAATCCA CATATGTCAA TCTATTAATG  
56251 TAAATCAATA TTGTATTCCT GGCTTTGATA TATAATATAA TTTTATAAGA  
56301 TATATAATCA TTGGGGGAAA CTGGATGAAG GATACAAGGG ACCTCCCTGT  
56351 ACTATCTTTG CAACCTCTTG TGTATATAAT TATAAAATAT ATAATGTATT  
56401 AAAATGTATA AAATAATATT TTAAGTATCA GATACTGATC TTTACTCAGT  
56451 ATATGAAGTG TTCTATCATA ACGTAACATG CTTTTCCTTT ATTTGTGGTA  
56501 TTTTAGTTTC AACTTAAAT ATAAATCACC TAAAGATCTA CGACAGTTCT  
56551 TTTGAAAAAA AATCTTGCTT TTAATTTCCC AGGAGTTTCA ACCTTAATCC  
56601 TCTCTTTAGT GTTCTTTTAT TTGGTAGTGA TAGGGACTAT CAAAGCTTCT  
56651 TACCATCAAA TACATTTACT GACTAAAAAT AGAAAAATAA TTTACATTGT  
56701 AAAAATGTAC AAATTGAATG ACAGTCAAAA GGTACAGGTA ATGAAGATAT

Fig. 2 (cont'd 31)

37/124

56751 GCATTAACAT CTACTTTTAA AAAAAAGTTT ATTAAAATTC TCTTTTAGAC  
 56801 TAATGCAGTA TCTGGGAATT TATATAAATA GATATGTATA TAAATGACTA  
 56851 TTAAACAATT TTAATGTCAG TTATATTTTA AACATTTTAA TAATATTGTT  
 56901 ATAACATATGG GGGTAAAAAT TTGTATATAT CTGAACATTT TTGTTCTTAA  
 56951 GGAAATAATC ATTTTACAT ATCCAGGAAT TTGAATTACT CTCAAGTCAC  
 57001 CTATTAATTA CAAGTCATTT TGAAGTCATT CATTTTCTTT GTGTTTGCTT  
 57051 TATAATGTCA TTTTAGATTT CATGCATCAT AATCAGCCAT CAAATAATTT  
 57101 AGTTAATACT TGATTTTCC TCAGTTGTAA GAAGTGCTGT GTTTAAATTT  
 57151 CATTGAGAAT GTTTCATTT ATCTGAATTA ATATCTGTTA ATGTATGTAA  
 57201 TATACACATA TTTTAAACAT GCATGTACTT AAATGATTA TAGGGACTTG  
 57251 GTAAATTAAC TTATTTATAG GATATTTTAA ATATAATCAA GGATTTTTTA  
 57301 AATCTACAGT TCCCATTTGA AAGTAAAAGT AAGTCTTTGT TTACTAGTTT  
 57351 GTTCACAGTA CAAGTAAACT TTCTACCTTT TGGTTAAATG TGAGTGCAGC  
 57401 CCCCACAGTG AGAAATTGTT ATATTAGAAC TCTAATAGCT ATAATTTATA  
 57451 GGGATGAATT TCAATGAGTT TGGTCTAAG AAATAATCTG TTGGTTTTAA  
 57501 CAACATTTTT AAGTATCAGA TATTCATCTT TACTCAGTAT GTGACATGTA  
 57551 CTCTCATAGC TTACGTGCTT TTCCTTTATT TGGGGTGTTT TTTATATATT  
 57601 AATTGGTATA TCGCATATTT AAAGTTGGCA TAATTACATT TATATGGACT  
 57651 CTAAACAATA ACTTGTATTT TAATTTTTAA ATTTGAAATG CATCTATGTC  
 57701 TCTGTAAAAA TGCATTTCTT TCCCTTTGCC CAAATGGGGT ATGGTAAGTC  
 57751 AAGAGAGTCT CTAGTTAGCT CACCTCTCAT TTGACTGGCA GAGTAAAGCC  
 57801 CTTGTTCACT AGAATGTGTG TTAAGCCTTC CCTCCCTTTT GTAAAGTTGT  
 57851 TCTGAACAGA GCTGCATAAA ACCACAGGTA AAGTGTTAAG CTGATTCTAC  
 57901 TAGCATGTCC TTAGAAAGGA GAGCGGTTAT ATTGGCAGGT CCTATTGCCT  
 57951 GCGGTTTCTG ATCAATAACT CACCAACAAA CAGAAAACAG AAGCCGCACA  
 58001 AGGAAAGGCA GAACTAAATA AATGGTAATA GCAAACAATA AGCCAGATAG  
 58051 CCTCTGGCCT CTGCCCCACA CCTTAAGGCA GCTGGGTCAG GTGGGATGCT  
 58101 TTTGTTTGTC TTTTAACGTA TTTTCTTTAC AAATCTCAGC CATTACATAA  
 58151 TTTGGAAATG GACACAAGGC TAGTTATTAC TAACATTTTT AAAGACATTA  
 58201 CTGAATGAAT GTGTAAGAAA ACAAAGGTC CTTTTTGCCT TTCAGCAGAT  
 58251 AAGTCTTTTA ACCAAAAATC TCTTGGGTAT TTTGAGATTG TGTCTACTT  
 58301 CTTTGCTTAT TTAATATTTT CATAAAATTT GCTAGTTACT CTGCTTTTTT  
 58351 TGCATCTCTT CTAAGAGAAA ACAATTGGTG CATATTATTA ATGAGAAACA  
 58401 CTTCAAGTGT TGGACAATTT TTTGTAGTGG AAAAGAAATG TGAACTTTA  
 58451 TGTTGCAGAA TCATTCTTGG TTCAACTAAC TACTAATTTT AAAACATAAA  
 58501 GTCTTAAATA TATATAAAGT TTATATGGGT AAATATATAT TACATATAAT

Fig. 2 (cont'd 32)

38/124

58551 ATATGTTTTA TATTTATACA TAATATACTA TATATTTATA CATGATATAC  
 58601 TAAATATTTT CCCATATAAA TAATAAAATG CTCTAGGCAT ATATGTGTGT  
 58651 GTGTGTATAT ATGTATATAT ATATATACCT TCATAACATA CATATATAAA  
 58701 ATACTATAT T ATATATACTC TAGGTATACA TATATGCCTA TATATGCACC  
 58751 TATATATTTA TATATTACTA TATAATATAT AGTATATATT ACTATATATA  
 58801 CTACTATATA TTACTATATA ATATATAGTA TATATATAGT ATATATTATA  
 58851 TAGTAATATA TTACTATATA ATATATAAAT ATATGTGTGT ATATATATAT  
 58901 ATGCCTAGAG TGTTTTAAAT TTGTCAGTGG GCTGTCTCTG TAATCTATAT  
 58951 GAAGAAATAA AATGTAGACG TTATGTATAA TGATATTTCA TCTTGTGTGT  
 59001 TGGCATCATA GTAATTCTCT TTACATATCT ATTCAGATTA CTTTTGCACC  
 59051 AGCCTAATAC ATTGTATGAT TCCAAAACCA AAGAGAGTAT GGATTGAAAT  
 59101 GATATTCCCT TTACTAATAC TCAGTCTTGT CTATTTTATT ACCTTTATAG  
 59151 ACTTCACCTA ACACAAGTCA GGGGATATTT ATCATCATAT TAATACAATT  
 59201 TTACTCTGAC CTTAAAATTA TGCAACTGCT AAAGGAAAAA TCAGAACCAA  
 59251 ATAAACTGTC ATTAACAACC CCCCTGAAAA TCCATATTTT TTAAGTCA  
 59301 TTTTATCAAG TCTCTCAGAC AAGATGTGAT ACCCTATAAG TTTAATCAGT  
 59351 TTTACTTTCC ATTTTCTCTT CATTAAGGTG ATAAAGATTA TCATTAGTAG  
 59401 AAAAATTTTC CCTTATTTGC CTCCTTTTCC ATTTACCCTA TTGAGTGAGA  
 59451 AATTTAGCCT CTCATAACTT CTAAAGTAGC AATGTTAATC TGATAAACTA  
 59501 AACCAAGGTG AGATAAATTT AAGACAATAT TTTTTTCTT CAACTTTTAA  
 59551 GTTCTGGCGT ACATGGGCAG GATATGCAGG TTTGTTACAT GGGTCAACAT  
 59601 ATGCCATAGT GATTTGCTGC ACAGATCAAC TCATCGCCTA GATATTAAGC  
 59651 CCACCATCCA TTAGCTATTC TTCCTGATTC TCTCCCTCCC CTAACCTCCA  
 59701 CTGACAGGCC CTAGTGTGTG TTGTTCCCCA CCATGTGCCC ACGTGTCTC  
 59751 ATCGTTCTAC TCCCACCTAT AAGTGAGAAG AAGTGGTGTT TGGTTTTCTC  
 59801 TTCCTGTGTT AGTTTGCTGA GGATAATGGC TTCCAGCTCC ATCCATGTCC  
 59851 CTCAAAGGA CATGACCTCA TTCCTTTTTA TAGCTGCATA GTATTCCATG  
 59901 GTGTATATGT ACCACATTTT CTTTATCCAG TTTATCATTG GCATTTGGGT  
 59951 TGATTTCATG TCTTTGCTAT TGTGACTAGT GCTGCAGTGA ACATAATGCA  
 60001 TGCAGGTATC TTTATAATAG AATTATTTAT ATTCCTTTGG GTATATACCC  
 60051 AGTAATGGGA TTACTGGGTC AATTTCTGCT TCCAGATCTT TGAGGAATCA  
 60101 TCACACTGTC TTCCACATTG GTTGAACATA TTTACTCTCC CACCAACAGT  
 60151 GTAAAAGCAT TCCTTTTTCT CTGAAACCTC TGCAGCACCT GTTATTTCTT  
 60201 GACTTTAATA ATCACCATTG TGAAGTGTG GAGATGGTAT CTCATTGTGG  
 60251 TTTTGATGTT ACCCTTTTTT TTATATGTTT GTTGGCTGCA TGACTGTCTT

Fig. 2 (cont'd 33)



39/124

60301 CTTGTAAAGTG TCTATTTCATA TCCTGTCTAT TCATGTCTTT GCCCACTTTT  
60351 TAATGGGGAA GTTTGTTTTT TACTTGCGCA TTTGTTGAAG TTCCTTGTAG  
60401 ACTCTAGATA TTAGACCTTT GTCAAATGGA TAGATTCCAC AAATGTTCTC  
60451 CCATTCTGCA GATTGTCTGT TCACTCTGAT GATAGTTTCT TTTGCTATGC  
60501 TGAAGGTCTT TAATTAGATC CTATTTGTCA ACTTTTGCTT TTGTTGCAAT  
60551 TGCTTTTGGA GTTTTTGTCA TAAAATCTTT GCCCTTACCT ATGTCTTGAA  
60601 TAATATTGCC CAGATTTTGT TCTAGGGTTT TTATAGTTTT TGGATTTTAC  
60651 TTGTAAGTCT TTAATCCATC TTGGGTTAAT TTTTGTATAA GGTATAAGGA  
60701 AGTGGTCCAG TTTTAATTTT CTGTATATGG CTAGTCAGTT CTACCAGCAC  
60751 CATTTATTAA TTGTTTTTTC AGTTTCCCCA TTGCTTGTTT TTGTCAGGTT  
60801 TGTGCAAGAT CAGATGGTTG TAGGTGTTTT TCACCTAACAT AATCATAACA  
60851 TACATTTTCAT TGAAAACAAC ACGACTCAA ATGTTCTTTA GTAACCAGTT  
60901 ATAAGTTTTT TTGTGCATAA TTACAACTG CCATTCTAAT CATAAACATT  
60951 TTGTGGTTAC TTATAGCTAG AAAATGTGAG TAATATAGTT TATACAGCAT  
61001 ACTCTTTACA ATCCCGATTT CTTTGTCAA CTTTAATTCA TATTAAATTG  
61051 ATAAAGTATA CACAAAGGGT AAAGGAGAGT AATTTCTTC AAGTTTCACA  
61101 TTTAAGGATT CATAGTAGAA TGATTAAACC TTACATTTCT CCACTATAAG  
61151 GAGAATTAAG ATGGAAATAT TGAGTAAAT CTTACATTTT ATTTAGTAAG  
61201 TGCTAATAAA GGGTTTCTGC CATAATTTTC CTTATTTTAA AAGAAAACAC  
61251 ACAATTTTAG TTTTAGGTTT TAGTAACCAA TTTTATGGGC ATAGTGGGAA  
61301 TATTTCTAAC AGGTTAAGT GAAGTGACCA TCATGGGCAT ATATATATAT  
61351 TTTAAATTCA CATATATGAA TACTATACAG TAAAACTAA CTTATGCTAC  
61401 ATACCACATG GATGAATCTC AAAACCCATG TAAAGCAAAA GAAAACCACA  
61451 AAAGAATCAT GCCATTTGAT TACACTTGGG TGGTTTTTAA AACAGGCATA  
61501 TCTAAACATA GTGCTTTTAA GTGTAAGCTT GGGTAGGAAA AACTATAAAG  
61551 AAAAGCAAGA AAATAATTAC CACAGAAGTT ATGTAGAGGT TATCTTTGGG  
61601 GAAGGAAGAG GGAATAATAA GAGAGGGACA AAGAAGAGCT TCTTGGTTCT  
61651 TGAAATGTCC TATTTCTTGA CTTGGCTGGT GAATGCATGA ATGTTCACTA  
61701 TGTGATAAGT CAGGGGGCTG TTTTCATTTT GTTCACTTTT ATATATGTGT  
61751 GGATTTTTC ACAGTTGAAA AGGTAAAGTT CAGGTGTGGT GGCTCACACC  
61801 TATAATCCCA GCCAACACTT TCGGGGGCCA AGGTGGGAAG AATTACTTGA  
61851 GGCTAGGAGT TGGAGAGTAA CCCAGGCAAC AGGTGAGGC ACTGTCTCTA  
61901 CAGAAAATGA AAAAAAAAAA AAAAAAGTAG CTGGGCATGT TGGTACATGC  
61951 CTATAGTTCT TGCTACTTGG GAGGCTGAGG CAAGAGGATC ACTTTAGCCC  
62001 AGGAGTTTAA GCCTGCAGTG AACTAGGGTT GTGGCACTGC ACTCCAGCCT  
62051 GGGTGGCAGC AAGACACTGA GTAAAAGAAT AAAATAAATA ATTAAAAGTT

Fig. 2 (cont'd 34)

40/124

62101 AAAATATAGG AAAAAATGAG CATAGCCTTA TGCTAATTTT TCAGTTACTA  
 62151 GGTCTGATAT CATCACATTC CTTGCTTGTC ATTGAAAATT TTTTAAACTA  
 62201 TGATACTTTT TTTTAGTGGT ATTTATCCAA TTAAATCTGC TAACAAATTT  
 62251 GGTGTATAAA TCTCAAGGGT AAGGGTATGT GGAGAGTGGG TGTGTTTGTG  
 62301 TGAGAGAGAG AGAGAGAAGA GGGGGAGGAG AAAAGAAGG AAGAGGGAAG  
 62351 GAATGGAAAA AGATAATAAA GAGTTGTTCT GATAGATTAA TCTTTAGTAG  
 62401 ATGTATTCCC TACAAATTGT TTTTCTCCAT ATTGCAGTGT CAGGTAAAGA  
 62451 AAGGCATCCC AGGATGAATT CAGAGCTAGG AACATGCACC TTTGTATCAT  
 62501 AATGCTAATG GAAGGAACAT GTACATTCTA ACTGTTACCA ATAATGGAAT  
 62551 ATATTTCCGT TATTAAGTAA TAAGCTTTAA TTCTTTGTAT TTTTGTGATC  
 62601 CATTTGATAG TAGGTGCCTC AGCATTTCCA CTCTGCTATA AGTACATGGA  
 62651 GATATATTTT ATTTAAGTCA TCTTATTCAT GTCTTTCAAA AAGAAATTCA  
 62701 TTTTGGCCA AGGATTTCCA AATTTTGCCC CATATATAGG TATAGTTTAT  
 62751 TATAGACTTC GTTTGCAAAA TATTAAATCC TTATATCCTT TTAGGGACAC  
 62801 AATAAAATTT TATAAGTTTG AGATAATGTA CTTGCAGTTC TACCTCAGGC  
 62851 CGTGGTGAGA GATTGAAGTG CCTCTTCATT TTAACATTTT GGGTTCAAGT  
 62901 TGTTCATATA GGGCATGCAA ATGGAAACTG GCCTATTTT TTAGCTTTAAT  
 62951 AAAATCGTCA AATACTTCTT AATCTTAAGA GTTATAGTTA TGTACTACAA  
 63001 TATGTATAAT TCTCTAATAT TTAACAACAA ACCTGAAAGC CACAAAAGCT  
 63051 TACTGTGAAA TAAAATGTGA TGGAAATATTA TTTCTAACTG GCTTACCTGT  
 63101 ATTTCTTCA TTGAAGGGAA TATGAAGTAG AAAAGCCCTT TTATTGAAAA  
 63151 GAGTTTGAA AGTAAAGATA ACTCTTTTCA ATTCAATTCT TTGTAAGTAG  
 63201 AAAAGAGTA AAGATAATGT TTAGCTGTCA GCAGATGTCT GACACTTGAT  
 63251 GGAGCGTATC ATTACAATAG AGCAGCTAAC AATATCTGCA AAGGTCATCA  
 63301 TGAAAGTATA AAAATGAGGA ATATTTGTCC ATTGACCATT TCAGTGACCT  
 63351 CTTTTTGGGC TTTAAGTCTA AAAATCTTGG CAGATCAGAA CTTTATATTC  
 63401 GGCATTTTGA GTGTCAAATC TCTACATGAT GTGCAAGTCA GAAGGAGTTA  
 63451 TTAAGTGCAA AATACCATCT TCTTTCAGAA GTTAAACTCA CATTAAATGC  
 63501 CAGGAGACTG AAACACTGAT TTTAAGAAGA CAAAGTTTAG AAAAGATGAA  
 63551 TGAAAATGTG TGTTAAAGAA GAGTCACCAG TCAGAGCTAA CTATGATAGT  
 63601 CATAGTATTT AAAGAGTTGG AACACATGAA ATTAAGCATT TTGTAAAATG  
 63651 AAGGCTTTTC ATCCATCCAC ATAAGATTCT GACATTTAAA CTATGTTTCT  
 63701 TCCATTCTGT TCACAGGCTC ACCTTGAGC AGCTTTTGAA AAGAGCTTAG  
 63751 GGAATATGAC TGGCCGATTG CAAAGTCTAA CTATGACAGC GGAACAAAAG  
 63801 GTATGTTTCA AAATTGCCAC TGGAGACTGA AAGAAGACAG CAAATTGCAT

Fig. 2 (cont'd 35)

41/ 124

63851 AGGATTCTTA AATAATACCT GAAGCTCCTT AAAAATAATA TTCCAGGCTG  
 63901 AGTGCAGAGG CTCATGCCTG TAATCTCACC ACTTTGGGAG ACCAAGGTGG  
 63951 GTGGATCACT TAAGGTCAGG AGTTCGAGAC CAGCCTGGCC AACGTGGTAA  
 64001 AATCCCATCT CTAATAAAAA CACACACAAA AAATTAGCTG GGCATGGTGG  
 64051 CGGGTACCTG TAATCCCAGC TACGCAGGAG GCTGAGGCAG GAGAATCACT  
 64101 TGAACCCAGG AGGCAGAGGA CGCAGTGAGC CAAGATCACA CCACTGCACT  
 64151 CCAGCCTGGG AGACAGAACA AAAAAAGAG TAATAATAAT AAAATAATAT  
 64201 TCAATTCTAT ACTAAATTAA AACAATGATA ATACCTTTCT TTTCAGATTT  
 64251 TAATTTAAAG ATTTTATCAG TTTACTCCAT ATTGGAACAC ACAAAGGCAA  
 64301 ACAAATCCT TGCTGGGCAG TCTATTAATT TACTTCTGGA TGGAACTAGT  
 64351 AAAAGAATAC TGAATGTTAA GAAAGAGAAA CAGTCACATA AGAGAATATT  
 64401 CTGGGGGCAA ACTGTTATGC AGTTGACAAG AATCACACTT TGATAAGAAC  
 64451 TTTCACAAAT ACATGGTCAC TAAATCCAGC TATAGGGCAT GGCTGTAGGC  
 64501 TAAGACACAC AGGAAGGATG CCTGGGACTC TGCCAAGTAA GGGACTTCAG  
 64551 GTTACAGCAG CTATGAAACA AAGGCCAATC CTGTGTAATT TTGAAATAAC  
 64601 AAGAACTAGT TGCCATCTAG GGATATCACC TTTGAAGAAA AGTCATTTGT  
 64651 TATATCAAAA TACTTAAAAT GAACCTAAAG GATTTTATGG TATGAAAGAA  
 64701 GGTATACCAA AAAGAAAGGA ACGGAGAATT TAGTTCACGA AGACAAATGT  
 64751 ATTAAAAAGG TCCATACTGC ATAGAAAGCC TGGTCACCTT TCCTGTGATG  
 64801 ACCAGTTAGC TTAATTCTCT GCTGTTAGTC CAGTGGCCTT AACTTCCTTG  
 64851 GATAGGTATC AGAGATAGGT GAAACCTATA GAATTCATG GAGTGTGTGT  
 64901 GTGTGTGTGT GTGCGTGCCT GTGTGTGTGT GTGTGTGTAT GAAAACTGTA  
 64951 AATGTGCATA AATGATCAGG TGTCCAGAGC TTTCATCTAA TTCTCAAAGA  
 65001 GACCCATTAT ATCAGAAGTT TTGGGTATTT TCAAGAATGC GTTCCTCTAT  
 65051 CTATCCATAG GAATGGCTTC AGTTTGTCT TTAGATTCTG TAAGTTATGT  
 65101 GATTAGCTTT ACAAAGTAG TATGTATTAC CAAATTTGT CACTTTACAA  
 65151 AAGTTTATTT TAAAAACAGA ATGAATAGTT CAATGAAATC AAAAGAGTAA  
 65201 ATCGAATATT CTTATAATTG CCAAGTATTA TTAGCACATT GTATTCTCTC  
 65251 TCATATTCTC CGTATACCCT GCCCGTGAGA GAGAATATTA TCCATTCCCTG  
 65301 GAAAATCTGT TCTAGCACAG CTAACAAACT CCTTTTGAAT CATAAATTTT  
 65351 CCTTTCTTTC CTCCCTCCCT CCCTCCTTCC CTCCCTTCCT TCCTTTTCC  
 65401 TTTTCTTTCC TTCCTTCCTG CCTCTTTTCT ATCCTTCCTT TCTCCTCCCT  
 65451 TACACCCTTT CTCCTTCTT TTCCCCCTCT GTCTCCCTCT CTTTCTTTT  
 65501 TGCTGCAGCT TGTCACCTCA CTATGTAATA TAAGAACCCA GCAAATAGAA  
 65551 TTAGAAGGCT TTTTAGAGCA GCTGACGGA AAGAATAAAA AACTGGCCC  
 65601 CCAGTATTCT TGAATGAGAA TTCTGGCTAT GTCTGTTAAA AGCTGGGTAA

Fig. 2 (cont'd 36)

42/124

65651 TCTTGAGCAA GTTTATCTAA CCTTCTTGA ACCTCAAATT CACCTTCTTA  
 65701 AAAGTGGGGA TGATAATGAC TACCTTGTAG GATCACCATG AGGAGTAAAT  
 65751 CAGATACTGT TATCATGTCA CATGCTAGGG GCTACCAAAA AATATTACCT  
 65801 TCCTTTACAT TTCTCTTTTT CCCTTGAAAA TTATAAGATA ACACCAAATT  
 65851 CCTCACTGGG CATATACCAA GCATATTGTT GGAAATGAGT GTTAGAATTT  
 65901 AAGTCTCAAT ATCTTTAATA AGTCAAAATT AATAGAATTT TTGTCCCTCA  
 65951 CCCAATATTT TCTTGAAGTC TGTATATCT GTAAGTGAAT TTTCTCATAG  
 66001 AACATACAG AGAATTTTCT CATATACATA TAGAAAAAAA TGTAGAGGTA  
 66051 TGTTAATGTA TAATGCCTAT GATTAATGCC TGAATATTTA AAAATAATTT  
 66101 CTATAACATA AGAGATTTTA TAATGTGTCT ACATAATCCT TAAAAAACA  
 66151 TTGCCAAAAT TATAAAATTT TCTCAGAAGA TATCAGAATG TCTCATATTG  
 66201 TCCTTATCAC TTTTTTAACT GAAAAATAAA TCACTTCTTT TTGAATTGCA  
 66251 AACTGTATAC ACACAACAAT CATGGTTAAC TAGTTTATTA ATTTGAGATT  
 66301 ATAACCTGCC TATTCTCAAA GTGATATTTA AAAGCCTATA AAATTATTTG  
 66351 CAATGTGAAA TGGTATAATT CAAAGACAGA ATCTAATTAA AACCAGTAGA  
 66401 ATAATGTATA TAACAATATA CCTCAGCCTA GATAATTACT ACTGCAAGGC  
 66451 ACTGAAATGA ATTGAATTTT AAGGAAGCTA TGGTACAAAG GGAGATTGTT  
 66501 AGGTGTGTTT TATTCTCATT TTCTGACCAG GAGAGCATAA TTTAGACTGA  
 66551 GGAGAAAACCT CTTTGGCACT AAATTCAAGG ACGAATTTAT TGCCAAGGTT  
 66601 TTTAAATTGG GGTCAATGAA TAACAAAAGA CAAATCACT GTTCAAATAG  
 66651 ACATTTCTCT AAAAGCTAAG GGCATAACAT TTAATCATAT TTCCTAAAG  
 66701 GCATTTCTTC AGGGAGCTGA GATAAAAGG TATATTGCTC TCTGGTGATT  
 66751 CAACAATCCT GAGAAAAGGC TTGTGAAGTA TAGAGCAGAG ATTCTTAAAC  
 66801 TCCCTTCCCC AAGTTATAAG TTTCAATTGT CTATATAGTC ATTCATCAAG  
 66851 TTTATATTGA ATTTGTGCTC TTCTAATGAC AAAACAGTAC AGACAATATA  
 66901 GATATAGAAT GATAGATATA GGTCTATATC TATAGACATA CCTATCTACT  
 66951 AGAACTCTAA AAGCATATTA TACATGTATG TAATATTCCT CATGGAGTTT  
 67001 ATATTTCTCA TATATATCTC ATATATATGT ATCTCTTTAT CATGGAGTTT  
 67051 ATATTTTAGG AGGTCACAGA TGATAATAAA AATATAATTA AAACAGGCCA  
 67101 GGTGTGGTGA CTCACACGTG TAATCCTAGC ACTTTGAAAG GCCAAGGCAG  
 67151 GTGGACTCCC TGAGATCAGG AGTTCAAGAC CAGCCTGGCC AACATAGTGA  
 67201 AACCCCATCT CTACTAGAAA CAAAAATTAG CCAGGCCTGG TGGTGGGCAC  
 67251 CTGTAGTCCC AGCTATTCTG GAGGTTGAGG CAGGAGAATC ACTTGAACCT  
 67301 GGGAGGTGGA GGTGTCAGTA AGCCGAGGTC ATGCCACTGC ACTCCAGCCT  
 67351 GGGCAACAGA GCAAGACTCT GTCTCAAAAA AAAATATATA TATATAATAT

Fig. 2 (cont'd 37)

43/ 124

67401 ATATAATATA TATATAAATA TATATATTAT ATAATATATA TATAAATTAC  
 67451 ATATTTATAA ATATGTAATT TATATATATA ATATATAATT AAAACATATA  
 67501 GGATTTTCAGG TGATGATAAG CACTACTGAA AAAAGTAAAG CTGAGAATGA  
 67551 GGATACTGAG AAGCTGGTTT GGAAGCTAAA ACACAAAGTA ACAAAGGCCA  
 67601 AGGTGGTTAC ATGTTCTTGA TTACATACTT TAAAAATGGA TAAACTAAAT  
 67651 TAAGACTCAG ATTCTAGTCT TTGGGCTTCA CAGTGTGATT TTCAGCAATC  
 67701 ACATGGCATT AATAGCCTGA AACTACATCA AAATTGTCAT TTGATTTATA  
 67751 GACCAAAATA ACTCCCTTGA ATAGAGAGGG ATTCACTCCT AACACTTTTC  
 67801 CTATTTCCAG ATGCCAAATA ACACGGAATC TCTTGCCAAA TTTGTGTGGC  
 67851 AGAACACTGG TTTTATATAC TTATAGCCTG GTAAGAAAGA AAAGACATGT  
 67901 ATGAATAACT TAGAAGGCAG AAAATTATCA TGCTATTAGA CTCAGTACAA  
 67951 TGTCATGTGC ATTCTCAAAG GAAACATCTG CAGAGGCAGG AGAATTGCTT  
 68001 GAACCTGGA GGTGAAGGTT GCACTGAGCT GAGATCATGC CACTGCACTC  
 68051 CAGCCTGGGT GACAGAGAGA GACTGCATCT CAAAAAATA AAAATTACAA  
 68101 AAATAAAAAA TAAAAAATAG TGATCAATCT GGCAGCATTT TCTGAAAGTT  
 68151 AAGCAGTATT CCCAATAGCT GCTAAAAGAA GACATGTTAT ATAATACTAA  
 68201 GTCTGTAAGT AGGTAAAAAT TAAGAGAATT GTTAATGTGC TTGCTGGGGA  
 68251 GTGAAATTAT CTCTAGGCAT TACCCTATAC CTAACCTAGG ACTCAGTAGA  
 68301 CTATGATATT GGCCTAGTTT GACCAAGAAT TTTATCCTGA TTTTCAGATCG  
 68351 TTTTCTCTTC ACCAGCACTT CTTCAACCAGG ATTATATGAA AAAAATTAAA  
 68401 CCTGATGCCC TGAGGCATCC ATTATATGTG CTGAAATAAC TTCTTTTCTC  
 68451 ACCATCTAGA ATGGTACTAG CTATGTACCA CTC'TGTCAG AATCAAGGAA  
 68501 ATTGCTACTC AAATCATTGT GCAGCTTAAT TTTCTCACAG AAGGCCAGTT  
 68551 GAGAAAGGCT CAACTTCTAG GAATCCAGCA AACTATATTT TTTATAAGTA  
 68601 ACATTTTAC AGAACTACTT CTAAATCCTT GTGTTCAAAT TTACTAAAGC  
 68651 TATATTCACA GCTAAATATT TCAGAATTTA AAATTTAAAA GACTTTCAA  
 68701 TTAGTTCCTT GTAGCTGTCA TGCCAAGGCA ATTAGAACAT ATGTTAAGGT  
 68751 ATGAGGGGTT TTTCTTGTTA GAAGGTCAGA GCAGGGCAGA GAAGTAGCCC  
 68801 CTTGTATGAG TGATGAAGCT CAGATATTGA CTCCTATGCT AACCATAAAG  
 68851 CCTAGTAGTT TGCTCATTTG TTACCTCTCT GAAACATTTT TTTGGGTGAC  
 68901 TACAAAACAG GAATTGAAAC CTTCAAAATA AGGGAATTG AAACCAAATC  
 68951 TTTGAAAATA GATAATGCTG CAACTAAAAA TTTAGTTGAA TAAGATTTTT  
 69001 ACATTAACTC TCCCTAATTT ACGTTATGAT ATTTGCCATC TAGAAGTGTT  
 69051 TTTAAAAAAT ATATTGCTGG AGTCAGATGA TGCATCCATT AATCTTTGGG  
 69101 GCATAGAATA ATGTGAATCT AAAATTTTCA AATTATTTAC ACTACTGGTA  
 69151 TTTGGTCAAT GTAATTTATT TGAAACTAGA TGCAATAGGG ATGGCCAGGT

Fig. 2 (cont'd 38)

44/124

69201 TATTTTCAGTA GAACAAC TAG CAAGACTTCA GATGCATGGT GGAGTGGGGA  
69251 AAGGAGGACC TGTTTAAGGA AACTAGAGCT GGAAGTGTG AGATTAAC TT  
69301 AGTGCCAATG TGAGGACCTA AAAAGCAGAT GTGGTGAAAA ATTTAAACAG  
69351 GCTTGCC TAG AAGGTCAAGT TAGTTGATGA CACTTGATGA GATTGTCCCA  
69401 AGCTTTGGGA TTCTCAACAA AGTCTTTGTT AGTGAGAAAT TTGGAAAGAG  
69451 ATCAGGTATA GTTAAGAAAC TGGGTTGGAA AGGCCACCAG GAAAGGCGAA  
69501 TATTC TGACA CAAAATTTGA TCATTTTATT TGGAAGCATT TCAAGCCTGA  
69551 CCTGAACGAA TTGTTTAGCC TCAGATACAT GCATAAAACT GTGAAAAGAG  
69601 ACATTGACTC AATTTAGCTT CTTTAACATG AGAAACTTTC GTGGAAAAC TT  
69651 AGAAC TTTAC AAGCTCAGCT GGTGTTGGGG GCATCATTAT CTTGAATAGC  
69701 TCACTGGAGG AAAATGAAAT CTTAGTTTGG TTCTCAGGTT TAAAAATATC  
69751 TATCATTTT GAAAAGTGTG AAGTAACAAA ATATGATCTG ATTATCTTAT  
69801 TCCTAAATC CTTTGCAGAA TTATCCCAGC CTCAATCTTC TCTTAGTAT  
69851 TTAATGAGAA TAAGAAACTG GAAATGACTG AATTGGAAGA GTAGACTTTA  
69901 AATCCATATC TTGATGGCAT ATACATTTT CAGTTTTTTT TCTAAATGAT  
69951 TAATGAGGAT TCTCAAAACT TGAGTATCTT CTATGTTTCC CTTCAACATA  
70001 AAGAAATTGT ATGAAAATAT TTTAAAAATT TCTAATGATT TTATAGTTAG  
70051 CTATCTTGGG AATTCATTT TAATCATGTA CCTCATCCAA ACTCCCCACT  
70101 ATGGACAAAA ATAAAAATAA AATTATTAGT TGCATCTGAA GGCCACATTA  
70151 CAATTTCTAT GCATTATAGA AACCTGAGAA AATGTATCTT AAAAAATAA  
70201 TGTGAACAAC TAACCATAAT TATGAAGAAG AAAAATGAAA ACTAGAAATA  
70251 AACTATTGAA AAATGTCTAT GTATCAGTTA AGTTTTTATT TAAAAATTCT  
70301 TTATGTTTAT CTCTATAATA CTATTGGGAA AGAGAGAAAG GAAAACCTGA  
70351 CTTTGTTCTC ATCCAAAGGA GGTGATTCCA CTGATTTAGC CAAAATAAGA  
70401 CTTCTGTTT ATAATAAATA ATAAAGTTTT TGATGTTTTT TATATGGTAC  
70451 CCCACTCACT AGGTGATCAG ACACCCTCCT GCAAAAAAAA AAAAAATACG  
70501 TATGCAATAA AGTTAAAGTT TTATGTTATT CTTCAAGGG GAGAAACATC  
70551 TGTTTAACAC AGACCAGAAT ATTTCAACAA AGTCATCCCA ATATTTATGG  
70601 AGATCATAAA TCAAGCGAAA AAATATATTC ATCAACAAC TAAACAACTA  
70651 CATTAAATAG TCTCAAAGCA CATTTTCACT TTTTTCTGA CAGGAAAACA  
70701 GGTTTCACAA GTGTGGAGAC ATTTTACCAT GGCTTTTAAC AGTGAGGAAG  
70751 GATGTTTAAA TAAAGGGAAA AATTATATGG AAAGCTCAGA GAAAAGAGAT  
70801 GGGTGTGGCT TGAGTGACAA GGTGAGAGCA GATCTCATTA ACTGAAATGA  
70851 GAGAGAAGGA AGGAATTTTG CAAATATGGA AAGATAACTA GTGCAAGTTT  
70901 GAACAGATTA TGTCAATCAA TGTAAGATTT GGCTATCTTT TTAATCAAAG

Fig. 2 (cont'd 39)

45/124

70951 AAGACTATGG AATATTTTAT AGGTGTTTGC TTATACTCAA AGTTTAAAG  
 71001 AAATAACAGT ATGAATTTGG TTGAACATA TTTTTCATA GATAGGATTC  
 71051 TCCCAAGTTA TATAGCATAT ATATTTCTTA ACTAGTTATT CTCCTTTTA  
 71101 CATATATTGT GCCACATTGA GTAACAATA ACCTGCTAAT AGCTATTGGT  
 71151 TTTTAAAAGA TAATTAATAT TAGAAAGTGA TCATTTTCT GTTTCATATT  
 71201 AAACATGATA TTCTGAAAA GCAACATTGC CTGAATGTTT TACATTTTAT  
 71251 CTTTTTGAAA ACAGGTTTTA TAAGAGATTT CTGTGAAAA GCTGAACGTT  
 71301 CTGACACTGA AATAAGTCAG CTAACCTCAA GCTAAGCTTA ATTTTTTGAC  
 71351 ACTGTTGGCA TGAGGTCTCA TTCCCAATTT TTTCATTTAA AGCCACAGGC  
 71401 AAATGTTTTA ACAGATTTTA ATCCGTAGTA CAAGCATTAT TGATCTTAAA  
 71451 TTTAAGGATA AAAACCTGAT TTAAATTAGA ATTTAATATG CATTCTAGTA  
 71501 TTTACGTTGT ATAATTAATA TTTACATTCC ATGATTCCAC TATGTACCAT  
 71551 TTATTTCTTT TTGAATAAAT TTCCAGTAGG AGCAGAATAA ATTTTCAGTG  
 71601 AATATTTTAT TTCTTGGGG ATATTTTTAA ATGGAAAATA TATTAAGTTT  
 71651 CGGTAAAATC TGTTGCTAAT TTGGCAGTGG ACAGAAATA AAAATTGGAG  
 71701 AGACTGAGTC ATTATGATGA ATTGGGTCTG ACTTTTGTC TGACACTGGA  
 71751 AATTTCCAC AAATATTATA TTCTTCTTTT ATAATAAATA TAGTCGAAAT  
 71801 GAATTGCAGT CAAGTATTTG AAGACCCATC TATAAATTTA GGCGGTACT  
 71851 GTTGATTTTT CATTATGAGA GATTCTTCCA CTCATAAGCT ACTAAAAGTA  
 71901 CATAAAGAAG GTCTGGTTGT TTGTTTTAAA TGTGACTGTT CTCTATCAGG  
 71951 AAAATGTCAG GTATCCGATG AAAATAGATA TATGAGGTGC CAGGTATCTA  
 72001 TTCCAAACTT GGATATCACT TCAATTAGCA TCATCTTTTT TTTTTTTAA  
 72051 AGTGTCTAAG GTTAGAATAG TCACCAGATA TTCCCATGTA TGAAGCAATT  
 72101 TTCTGCAAAG GCCGCTGTGG ATGATCTTTT TAAAATATAT ATTTCTGGAG  
 72151 ACATTGAGTA AAGAGAAATT ATTTACCAGA GAATGAAGAA CCGAGGCCCC  
 72201 ATTCTTTGGC TTTCTGCCAA AGATGCTGAA GGCAGTGATG AATGACAAAT  
 72251 ACATTACCAA GGAATTCTCC CTCTAAGAGG CTGACAAAGA TCTGATTTTT  
 72301 AGGATTATAT TACCACCAAG AAGATACCCC TTGTCACTGA GCTTCTAATG  
 72351 GAAATATGGT CTATACTGAA ACAATTCTCA GTTCTTTTTT TTTCTATCTT  
 72401 TTTTTGAGTT ATTTTATCTT CCAAAAATGA GTTATTTCTG ATAAAATAAT  
 72451 TCACTTAAAT AATTATGAAA GTTCAAATTT GTGCAAATAT TTTTATTGGG  
 72501 ACATCTTAAA ATTACTCTAA ATTCAAAAAG AAAATATATG CTTTATTAAA  
 72551 ATTTGATCTG TAAGCTGCTT TGTTTGTAAT TTAACATTA TATAAAAATT  
 72601 GTATAATACA TATATTTTAT TTACTTTATT CCTGTGTTGC TTTGGCTTGG  
 72651 TGAGACTAGG TCTCCACATT AGGAGTTTTA CTGAATGAAA AAGTATCAGA  
 72701 ATGTAACATG ACTTTGATAT GGCATCAGAA TTTAATAAGA TGACATTTAA

Fig. 2 (cont'd 40)

46/ 124

72751 TAGGAATTAG GGGTAAGTTC CAGGTTTAC ACTTAAATAC AAATAATCAA  
 72801 TTTTGCAGGC ACAAATACT TCAAACAAAA TCTGAAATCA TTCATTTGAC  
 72851 AAAACTTCAG GTTTGCAGTT GACAATAAAT ACAATACAAT GCAACAGTGC  
 72901 AATAGTGATA TCTAAATATC TAATGTAATC ATAGGTAATA TTAGTAAGTG  
 72951 TGTTATCTGA AATGAGTGGT GTGATATCCT GCTTTACTTT GTACTGGTGA  
 73001 GTTCTGGGTG CCACCTTTGA AAGGAATAAA GACTATTCAT ATCTCTTTTA  
 73051 TAAGACAATA AGAAAAACAA ACAACAAAC AAACAAAAAA CCACCTCCTT  
 73101 TACTTTAGCT GAGAAAGAAG TTATTAGGTA CAGCTTGACA AGTTCAGCTA  
 73151 AGCATCCAAA TCTTCCAGGA GGTTGTTACT ACATAAAATC AAACCTTTTT  
 73201 AATTCAACTA TGAGCAGGGA GATTTTATTT TTCTTTCGGG TACTAAAGCT  
 73251 TCCAAACTCT GTTTATTCCA CAGGAATCTG AACTTATAGA ACTAAGAGAA  
 73301 ACCATTGAAA TGCTGAAGGC TCAGAATTCT GCTGCCCAGG CGGCTATTCA  
 73351 GGGAGCACTG AATGGTCCAG ACCATCCTCC CAAAGGTATA TTTAGAAATC  
 73401 ATTTCATTTT CACCCAATAT AATAGGCATC TATTTTATTT ATTAATTACA  
 73451 GTAGAACTGC ATTTACTCAG TGTCACTGTG CATTATTAAT ACATACTAGT  
 73501 TGTATTAATA GTTGTATTAA TACATACTAG TAGTATTAAT ACATACTACG  
 73551 TTGGTATTAA TGTGATCAGA ATCCTAGAAT TTTAGAACAG TGACTTCCAT  
 73601 TATCAGATAA TTTTAAACT GATCTTAAGA AATTGGTTC TATAGTTGTA  
 73651 TACACATCTC TCTACTGAT TCAGTGGAGA TGGAGATGGA GTGGTTGGTT  
 73701 AATACATGCA TATCTGACTT CAGGCAAAAC AAACCCATTA ATGAGTATGA  
 73751 TAATCTAGAT CTGTATTTAA AAATGAAATA GTCAATATGA TGATATAGTA  
 73801 AGCAGTGGGC ATTGGGAACA ACTTTTCCTG GATGGAGGCT ATAAAAAGGT  
 73851 ACATTTCTCG TAGATAATTT TGAAACAATA AAAACAACGG GTGAAAGGTA  
 73901 GCTCTGTTTT AAATTATTC TATGCTTAAG CAATTCTAAA CAATGAAAGG  
 73951 GGTATTTCTG CCACTGCCCC TACCCCTGGG TTCACCACTG AAGAAATGCT  
 74001 CATTATTAAT ATCGTGTCAT TTTTTCCTT TACATTGGTT CTATTTACTC  
 74051 ATTTCTTGAC ACTTTTCAAT GGCCTCAGT GAGCTCAGCT CTTTCCCAGC  
 74101 TTAAAAATC CTGTCCTAAA ACATGAATGC CTTATTATCT CTCTTTTCAT  
 74151 TTCCAGAAGA ATTCTGAGAA AAATTTTATG AAGTCTTCA ATGTCTTCAG  
 74201 CCATCTTTAG ACCACTGGAG TGTAGCTCCT TTTCCCTCCA CTCCACCAAA  
 74251 ACAATGCTCT CCAGGATCAG CAGAACTTA CATGACACTA AATTCAGTAA  
 74301 AACGTTTATA ATTCTTATTG TATTAGACAG ACATGGAAAC AGCATTTGAT  
 74351 GCTGATATTC ATTTCTTCCT ATGTGAAACA TCCGGTTTTT CTAATGTTG  
 74401 TGACATCATA CATTCTTGGT TTTTCTCTG TTCCTTTGAA ATATTTTTTC  
 74451 AATATTTCTT TTGTAAATTC ACTCTTTTGT ATCCATTTGT TAATGTTGA

Fig. 2 (cont'd 41)



47/124

74501 TATCCTAAGC TCTCTTCCAT TATGATTCTA TGCATCCTAT TTAAAATATA  
 74551 TAGAAAATCA TCTCATACTC TAGCTGTAAT TTTTATTAAT GTGCTAATAG  
 74601 CTAATAACTG TCAAATCTAG GTCTCCAGGC CAGGCTCTGT ATATCCAGCT  
 74651 ACCAAGAGAG AACTCCACGT GGATATCTTT GGATGTCTGT TTTGCATCTT  
 74701 AAACCTAACT TCTCCAAATT TGCAC TTGTC TTCTGTCTCA GACCTGCTGC  
 74751 TCCTTCAGTG CTCTTTGCCT CAGTAGATAG CACCACCATC CTTCCATTTA  
 74801 GCCAGAAATC TAAGTATTCT TCATAACTCC TCCTCTCCTC ATTGAATAAA  
 74851 TTACCAAGAT CCGTTGATCC CATTCCTTAA ATATCTCTTG GATCTGTTAA  
 74901 CTTTTCTCTG ATTTTACTCT TGCCATCCAT CACCTCTCTC CTGAACCATG  
 74951 ACCACAAACC CCTAAATAGC CTTCTCTCTC TTAATCTTAT CCTGCTTTAC  
 75001 ACCAGTCTTC ACGCTGAAGC CAGAATAGTC ATTAAGAAAC ACATCTACAG  
 75051 GTATCCCATT CATTCGCTTT AGAATGGAAT ACAGACTCCT CAGCATGACA  
 75101 TAATCTCTCT TCACCAGCTT CATTTATTCA ACAAATATTT ATTCATAACC  
 75151 AATTAAGTGC CAGATGATGC ACATATAGAC TTCTTGTCTT GTTGTTGCAT  
 75201 TGCATATTCC ATATTTTCTC TATCCTGAAT TGTTTTCAAT TATTCATAAG  
 75251 TTCTTTATGA ATTGTGTTCA TTCCATTTGG AATATTCTAC CTTGTTTGAT  
 75301 CAGCATAAAG ACTTTTCGAG AACTGCAGC AGCAGTGAAC CTAAATATGT  
 75351 TTCTTGACC CCTACATTGA ATGACACCCC CTGTGATATG TTTCTGGAAG  
 75401 CAGCAATACT TCCCTTCTTA AAATTACATT ATACTTTGGG GCTTTTATTT  
 75451 AAGGTATGTC TTTCTTGATT TACAATAGTA GAGCTTGTTT TTTTACCCTT  
 75501 TTGAAAGACA TCAAGATGCC CATGATGATG TCTTGCATGT AACAGGGGTT  
 75551 TATTTGAATT TTTAAAAGAA GAATAAGTA ATTTTAAAT GAATTTCAAT  
 75601 TTAAATTTTA GGAAAACAAT TATATAAGT GAGATATGCT TAAATTGAAG  
 75651 GACAAAGTAG TTCTGTAGGG GCTACTTCTT TCAAGACTTT AGCAACTTTC  
 75701 CATGTGGGGG AGTGATTTAT GTGATGCATG GAAATTTACT GCATATTTAA  
 75751 AGCTTATCTT AGAGCTATAA TAAAGCAGCT TATGTTCTAA ATCTTCATGT  
 75801 CGTAAATAGG TCCAGAAGGG ATTTAAAAAG CCTTAATCCT TACTTTAACA  
 75851 CAGCACAAGT CACTGAAGTG AAAGTGCTG AAAGGATTC TTTTATGTTA  
 75901 GGCAACAGGT AGCTGAATAT ATCTACAGAA ATTGAAAAAT TGAATTCTT  
 75951 TTGCTCAGAA ATGTGGGAGG GGTGGAGCTT AAGGTAAAAA ATAACAGTTA  
 76001 ATATCTAAAT TGATCAAGAA ATATGAAAAA ATAATTTGCT AGGTTTTTAA  
 76051 ACTAACAAAA ACCATGGTTA TAAAGGTTTG AATATATATA GGATAGTTAG  
 76101 ATTGTATTTT TGTAATATTA AAAGTCAGCA TTAAATTTAA TGAACACAAA  
 76151 GTGATTCTTA TCACATTGAC CATTGACATT ACATGGAAAA AATAGTCAGT  
 76201 TGGACTAATT ATGTGTCTTT CCATGGGTTA TTAAGGTAAT TGTATGGCAT  
 76251 ATAAATTTAT ACTGGAAATC ACATTGAAAT TCACTTTTAG AGGCCCTTAA

Fig. 2 (cont'd 42)

48/ 124

76301 AATATTTCTG TAATATATAT TTTTAACATA TGATCTTAAA AGATATATTT  
 76351 GGAATGACAC AACAGTTTTA TAGACAGGCC TGACTATCAC ACAACCACAC  
 76401 ACCAATTTGT GAATGTGTTT CTATTTCTCT TAAATTAATG CATCACATTC  
 76451 ATTAACAAAG TTTGATAAAT GACTATAGTC TATAATAAAA TATTTTGTGTT  
 76501 TACAAACATA TTTAAACACC TGCTATTAAG TATAGGCATT ATCAGATCTT  
 76551 AAAATACAAA GATTTAAAAA ATTACCCTGT GGTCATGGAG CTCACAATCC  
 76601 ACTGCAAAAA TAATGTTTGT GATAAGAAAT TTGAAAGTTG AAGGTAATAG  
 76651 AAAATTTTAC CTTTATTTT CAAAATGTAC CATTGCTTTC TAAGTCACTA  
 76701 CTTCTGTGTA AATATGGAAT TGTTTTTCCT TAAGATATAC CAAATATAGT  
 76751 TGGATAACGC ATGTATTAAA ATTCTGTCAG CACTAAGTTG TTTTTTAGAC  
 76801 ATAGTGATAG GCAAACATAG TTATATTGAA TGAAAAATTA GAATCAAATT  
 76851 TATTAAACAC TGTGTACTGA TTGATACCAC ATGCCATATG CTTGTATAGC  
 76901 AATACAAGGT TTGGAATTTA TAATGGTAAA CAAAATAGAT ACGGTCCTTG  
 76951 TCTCCATAGA ACTTTTAGTC TAGTGGGAGA GCAGAAGGTA AAGGAATGTA  
 77001 TGTGATCATT GGTGAAGCTG AACATGTATA CCCAAACAGT TATAAGTTCC  
 77051 AAGATGGACA ATAATGGGTG CCATAGGGAA GGAGGGTACC AAGGAACCTA  
 77101 CTGGAGGTTA CATAGGGAAG ATTATTCCAA GGTAAGTAATA TTTAAGTGAA  
 77151 TATCCAAGGA ATAATTGTCA ATCACTTTAT AAGTACTGAG GGAGGAGTAT  
 77201 TTCAAAAGAG CTTTGAGGCG GAAAATAAAT TAGTTCCTTT ATGGAAGTAA  
 77251 TGTAAGGAAA ATACTAAGCA AACATGTAAT AAGAAGAACA CGGTTGATGA  
 77301 GTTAAGAACT GACAAGATTA CTGAAGGATT GTAGGCCATA TTTAGAAGTT  
 77351 GGATTTTTTA TCTATTCTTA TTAAAGTGAG AAGTTATTGA AAGGTCTTAA  
 77401 GTGGGGGACT GATGATGAAG TTTGCCTTTT AAAAAAGATT TTTCTAGCTA  
 77451 TTGTTTATAG AATGGTTTGA AGATGAATAA GTCCAATAGC TATACTTGCT  
 77501 GTAAAGGTTA TGTTGGTAGC TTGAAGTGGG GCAGTGGTGA CACAGAGGAT  
 77551 GGGAGATGGA AAATGACGAG TGAACAAACA CATACCTGAA AATTTAAGTT  
 77601 TAAAAATAGA CCTCTCCATT AATTCAGATT GCTGATATTC ATTCGGTTAG  
 77651 CCATTCTTTA CTGAACTTTA TGATGCCCCA TATACTGAAT TAAATACTTA  
 77701 CAAGCACTAA AAAAGAAATT GTTAGGGAAC AGTAAAATGC ATTTCTTCA  
 77751 TTTCACAATA TTATTAATAT TATGGCTTTG CTAATCTTTA TTGGTGAATG  
 77801 CAGTCATAAT TGAAGGTAAC TGATACTTCC AAGGACTACT TTTGACCTAG  
 77851 GATTACTATC TTTTAAAAA TTTAGTATTA AAGAAGTCAA ACACAATTTA  
 77901 TTAATCTCGG ATATAATAAA AATTCTGAAA TACTTTAATA CTTTGTGCTT  
 77951 TTCTATTTGT GAAAGTTAAT TATTAGGAAC GAGCTAGCAA ATGCTACTTC  
 78001 TTTTTCAAAA AGCTAATGGC CAATCACAGC AAAAATTTAA AGCACTAAGA

Fig. 2 (cont'd 43)

49/124

78051 AATACCTACA CATATTCTTC TATTGCCCAT TTATATGACT TCCATAATAG  
78101 TTGATTAAAG GATACCGGAT TCCTTTATTG TTGAATTAAA ACCTCCTACA  
78151 TGAAAACCTT GATTTAGGTT TAGAAGTTGG TAATGTTTTG GCATGCAAAA  
78201 CCAGTTAATG TTCTCATCAT TACTTTTTTAA AACAATGTTA AGAGATGAAT  
78251 TCTAGGGATT ATAAAAAAA AAAAGCTGTA TGTGTTTCTT CCTATAAAAT  
78301 TTTTCAGCAT GATTGCCTCA GTAGAAAAAT TAAGGGACTT ATTGATATAT  
78351 ATGTATATGA AGGTGAGGAT ACACATATAC ACACACACAT ATATATGTAG  
78401 GTAAATACAT ATATTACATG TCTATCAATC CATACTACT CATTATTAT  
78451 ACGTTTTGAA AGCAACCAGT TATAGTTTTG TTGCCATGGA TCATTTTTAC  
78501 TATTCAGTAA ATCAGTCAAT TGAAGAGGCT TGATTTTATG GTATTAGTTT  
78551 TTTGGAAACT GTCAGCTTTA TAGTAAATTT TGACATCTTA CAACTTCCAC  
78601 TGAGATTTTT TTGCTTGACT AATCTGCCTT GATGCCAATA AGTATATTAA  
78651 CGGAAATGGA CTAAAAGCAA ATGTGACTTG AAGCACAATT TTGTAAATTT  
78701 TCTTAGTGTC TCAGTAATAC TTAATACTAG TGCATTTTAG GTAGGAAAAT  
78751 TTTCAGTTTG TTTTATTTTA AATAACTATA AATCTTATAG TTGCTTGAT  
78801 AAAAGAAACA GATACCTTTA ACATGATTAA ATATCAAATG CTATTCTCTT  
78851 CAAAATATCT TAACTAAAGA AGCACTGCCT GCTCTTAGAA GTTAAGCAAG  
78901 GCCATACCAT ATGCTGCGTA CATGGCTTTT AACACAATGG ATATTAGAAA  
78951 CAGCCTAAGG CTGAGCCTGG CTCCACTATT TTTCAGCTAT GTGACCATGT  
79001 GAAAGTTACA TTTAGTAATT AAACTCATTT CAGTAGTTTG CTTTAAGAAT  
79051 AAAATTAGGT ACTCCGGGGG CATATCAAGC ATATTGTAAA ACCTAGTTTG  
79101 ATTATTATTT GTTATTGGTA TTACTATTAC TATTCTATAA TAAGTCATGG  
79151 GCAGGCAGTA GGGGTACATT GGAAGAATTG CACTGTCTTA AATATGTCCT  
79201 CTGTTTAACT CACAACTCA GTCTACCTAG GCTTTCTTTG GAGGATCTGC  
79251 CTTTCATTGG CTGTTTGACT TTGGCCAAGT TACTTAACTT CTTTTCACCT  
79301 CAGTTTCCTC ATCTGTGAGA TTATGTGCTT ACATGACTTC AGGTTTGTGTT  
79351 TTGGCTCTAA TATGGTATGA TTCTATGAAA TGGAAAGTTA ATACATTTGG  
79401 CTCTAGTAAC TGTATTTGAA GCACAAATAT TAAAAAGCAC AATTAATTCT  
79451 CATTCTGAGT TTCCATTTAC TCTTTTAAAT TAATCATTTCA GAATAAATCA  
79501 TTTTGAAGA GCTGCTTGAT CCAGGTATTC AGTAGAAATC ACTAGCATAG  
79551 CATTTAATTT TAGACAAAAC TGAGAACTCA TTAAACTGCC AGGGCTATGG  
79601 ACTTATATGA GATTCTCATT AAATCTTAAT GTAGATAACT CAGTTAATTA  
79651 AAACAAATAT GGTTGTACTT TATTAACTT CTAAAGTCAA AACTGCATTG  
79701 AAATTATCTG TACAAAGCCT TGTTGACCTT TATTAGAGAA CTGCCTCTCA  
79751 AAAGACCTAA AAGACTTATT TGTTGAGATC GAGACTCTTC ATGAGCCAAT  
79801 GTGATACTCT CCCTCTATTG CTAGATCTTC GCATCAGAAG ACAGCATTC

50/124

79851 TCTGAAAGTG TTTCTAGTAT CAACAGTGCC ACAAGCCATT CCAGTATTGG  
79901 CAGTGGTAAT GATGCCGACT CCAAGAAGAA GAAAAAGAAA AACTGGGTAA  
79951 GTTACCATCC TTCATCTAAT TCAGAAGCTT ATTAATGCAT AATGTGTTAG  
80001 GCCTTTTCT TGGGGCTTT AGTGATCTGC AGTAGTTTAC AAAGGGTCCC  
80051 ATTCAAGCTA CTGAGACCTC AAATGCTGCA CTCATCACCA AAATTGGAGT  
80101 GGCATGTACT GAAAAGCATA CATTTTAATG TTGGGACTAA ACTTGGGTTT  
80151 GAATCACCAC TATATCTAGA CCTTTTGAGG GGCCTGAATT TTCTAACCAA  
80201 TAAAAAGACA GTTAATAGCA ACTATATTTA TTTGTGAATA TCATTTATTC  
80251 ACAGATGTTA TCTAATTTTT CTATAGTATA ACTATACAAA CTATGTAGTA  
80301 TAACTATAGA GTTATACTAA AGAAAAATAA GATAACATCT GTGAATAAAT  
80351 GGCTTAAAAT AGGGGTTTAT TGTGGGCATA GAGATGAAGG AAAAGTGAAA  
80401 AAATGATGAT GATGGTGATG ATGATGGTGA TAGTGGTCTT GGAGGAAAAG  
80451 GAGAATGGGA GTTAATAAAG GGAAAGAATA AACAATGAAA CTCTCATTC  
80501 ACCTTTGGAA TCGACAGGGC TTACCGTGTG AATAGTTTCA CCCTAAAAGA  
80551 AATCAACCAC ATTAGTGTCT GCTTGATGTT TTTAACCAAG AGAATATAGC  
80601 AGAAATATAG AAATGCACTT TAACAGAACT GTACCTTAAG TTTGCTAGTG  
80651 ATATAATTTA TGATATTGAT CAATAGCTAA ATAGCCCAGG GGAAGATACT  
80701 GTTACTGCGA AAAATTTAAA AACAATGGAG TCAATGATTT CTTTAAATAC  
80751 CAAAAAATA ATGTAGATTT TGAGTAAATA CAACTCTTGA TGAAATCCAG  
80801 ACATAATTAT CAGAGGATTT TACTGGAGTG CTTTCTACAA ATAATGAAAG  
80851 AAATATCTTT TTATCTTAAA AAATGTTTAT ACAGGTAATA TTTTAAAATA  
80901 CTGATCAGCC TTCATTCCCT TGATTTGTAA TTCCCACTC TTTCATGTTT  
80951 CTGCAAGGTG AACTCTAGAG GAAGTGAGGT GAAXATAAAC CGTGGACAAT  
81001 TTGGCATGGA TTTATAAAAA AACCCTACCT TGGCATGAAT GCTATCCATT  
81051 TTGGCAGTAG GCTTTTATAC CTTTAAAAC AGATTACCTT GTATGTCTTT  
81101 TCTTTGTGTC TTTTCATTTT AATCTCAAA TTTAAAGAGA TGTAACCA  
81151 CTTTCTGAAT AGAGCTGTAG GGGATACCAA TTCTGGTTTT GAGTAGTCTG  
81201 GGGTTGAAAA ATTTGAATAG AAAAATCACA ATTAATGAAG TGTTAGGTGA  
81251 ATTTGATTTT ATTTTGCTTT TTAAGTTTGT ACTGTCAGCA GGACATGACT  
81301 TGATTGTAGC GCTAAAGTGG CCATTTAAAA CAAATGCCT TGAAGAGAGA  
81351 AGCATTGGGA ATGGAGATC

Fig. 2 (cont'd 45)

51/124

## Human genomic sequence

1 GAATTCCTGG TGGAGAACAG CACATGTACA GATGGGGTGA GAACAGCATA  
51 CGTACAGGTA GGGGTAAGCT GGTGCTATAT GAGAAAGCAT GGAATAAGTT  
101 ATTAAGTTTG ACCTGCTTGG GAACTGAGGG GCAGGTGTGA GGGATGAAGC  
151 AGGAGTAGGT AGGGGCTAGA TCACAAAAGA TCTATGCCAG TGTTCCTCAC  
201 AGTGTGATTC CCAGCCCAGT AGCATGATAT CACTTGGGAT CTTGTTAGAA  
251 ATACAAATTC TTATACATCA CCCTGGACTA GACCACCTGA ATAAGAAAAG  
301 TTGGGCATGA GGCCTACAAA TTTTAAATAA AGTCATACAG GTGATTGCAA  
351 TGCATGCTAA AGTTTGAGAA ACACCTCTTG CTGTGGTTTG AATATTTGTG  
401 TCCTTCCAAA ATTCATGTAG AAACCATCTC CAATGTTATA GTATTAAGAG  
451 GAGGGACCTT TGGGAGCTGA TCAGATCATG AAGTCTCCTT TCTTATAAAG  
501 GGGATTAAAA GCCTTGGCCC TTTTACCCTT TGTCCATGTA AGGACACAGT  
551 GTTGAAGCA GGGACTGGGT TCTCACCAGA AACAGAACCT GCCAGCCTCT  
601 TGGTCTTGGA CTCTCTCAGCC TCCACAATTG TGAGAAATAA GTTCTGTGTG  
651 TTTATAAGTT AACCACTCTC AGGTATTTTG TAATGGCAGC ACAAAGGGGC  
701 TAAGAACTG TTCTATGCCC TAACAAGAAA TGTGGTCACT TTCCTGAAGG  
751 AAATGGGGAT ATATATAAAG ATGTTATATA AGACTCGTAA TATTTATTTG  
801 GAAGGCTTGC TCTGCAAGCA AGGTGGAAGA GCAACATGAA GGAAGCGTGG  
851 TGGAGGTGAG AGGACTGGAG GTTAAGTTGG TAGGGAGATA CAGGAAAGAA  
901 GCTTATGACA CTTGAGTTAA AATGTAGCAT CCTTCCTATG TGTAGGGCTC  
951 ATAAAAATGT ATAGTCTAAG ATAGAACACA GAATACTCTA TGAATCCTGC  
1001 CCACAAGGTG TTGGTAATCT AGATTCACCT TTTTTTCTG ATAATGCCAT  
1051 CCATATGTAT GGAGCGTCTA CTACTGTATG CCAGAGTGAC TCTGGAATCG  
1101 GTTTGGTTGA TCTAGACAAG ACCATAAGGA GAGTCCCCTT ACTACCTCTT  
1151 CTCCAGGGGA GGGATTCAAG TTGAACTAGT ACTTCAGAGA CTGTTTAGTA  
1201 ATATCATGCA TGAAAGGTGA TGGTTAGGAC AGAAAAATAA ATGGATTGCA  
1251 TCATAATTC TCAGGTCTCT CAAATATGTG GTGGTCTCAA ACCATGTGAA  
1301 TTGGTCTGCA CATCCTGTTT GGGTTGCGTG TCAGCAGTTG AGATCTGAGC  
1351 CTTATTTGTA ACAGTGAAAC AGTGAGAGAC CTGCCCTTCA AGAGCTGTTT  
1401 TTCAGCTAGG AATAGAAAAG GGCCAGGCTA GACTCCTCTT TCTGCTGGAT  
1451 CTTGCTTCTT CTCAGCAATA GAAGTAGACC TGCCTTCCTA GCTGTAGAGA  
1501 AAAGGTGCCG GTAGGCGGGC AGGTGAGCCT GTGGATAATC CTGGAGTAAA  
1551 GGTTCAATAG ACCTTCAAGT CTATCCTACA GGATTCGGAG TGAGGGGAGA  
1601 GAAAAGGAGA CGCTTCTCTG GCTGAGAGAG GAAGAGAAAA AAAAATCCCA  
1651 GATATCTGAC AGCTATATCT TCCCATCACC ACCTTCCTCT AAACCCATGC  
1701 CTCTCTGTTT AGTAGGACAT AAAATGAAGA GTGACCCACC CCCCACCCCC

Fig. 3

52/124

1751 AGCCCATCCC CCGTTTGTAG GTGTGCTTTC AATGAAAATA AGTCGGTGTT  
 1801 CATGGACGGA AACTAGAGCA GCTGAAAATA GATGCAAGAC TTGTTGAGCA  
 1851 TACAAATCAT TTCCCCCTTA GTCTCCAAGG GAGGAAAAAA AATCCCTCTT  
 1901 ACTCTCCTTG CAGCCTGTGT TCTGCATTCT GGAGAGGAAG CTGAGGCTGG  
 1951 TCCTCAGGCG CTCCTCCCGC CGTTCCTCGCA GGAAACTTTT CTCGCAGGGC  
 2001 CCGCTCCGTC CATCCCGCGC GGTTCCAAGA CCGTGGGCCT CCCGTGGGCT  
 2051 CCTCTCCTGG GCAAGGGCCC AGACCCCGCG ACGCGCCTGT CTCTTTAAAT  
 2101 TCCAGCTGCG CGGCTGGGAA ACAGCGCCAC TCGCCGCCCA GGCCGGCTGG  
 2151 AGGCTGAAGA GCGAGCTCGC GCTTTCGCTC CCGGCTGCGC GCCGCGGAGA  
 2201 GCTGGGCTCG GCCCGCGGGC TGCTAGGTGG CGGCGGCGCG GGGCGGGGAG  
 2251 GCGCGGCCCC GCGGAGGAGG GAAGAAAGAG CGAGCCGGGC CGGGAGAGGC  
 2301 GCCGCGCCCC GTCCCGCGCC CGGTCCCGCA CCCGCTCTCA GCGGCCCAAG  
 2351 CAGTTTCTTT CTGGGTGACA AGAATGTGCC TCGGTTGGTT TTTCTTTTTT  
 2401 TTCTCCATCT CCTAAGACG ATTTCCATAG TAACCTGATC AAGTGGCTCA  
 2451 AAATCGCAA CCTGAGGATT TCCGCGGCC GCCGCAAGA CCTCGGCCAG  
 2501 GTAACGCTGC GATCTCCTCC TCTTCCATTG CAAACCGCTG CGCTCCTTGC  
 2551 AAAGTTCTTT TTGTGGAAAA TCGCCAGCC CAAGGGAGCC CGGGGTATTT  
 2601 GCAACAGCGT GTTCATTTCC AGGTGCCTGT CACGGGTCTC CTCCCTGCTG  
 2651 CTTCTCCAGG ACCCATGATG AGATTATTTT TAAAAATTGT TTTTGGTCGT  
 2701 CTCCCCCGCC CCCTCCCCTT CTTTATTTTT TTCTCTTCG CTGCACTCTT  
 2751 CTCGGCTTTT CCCCTGACAC TACTGATGGG GTGCGGGGG GACGTCGGGG  
 2801 ATGGGGGTGG CCAGCGCGGT CCTGGGAGTG GCGGGTTCGG ATGGGCTGGC  
 2851 TGCGGTGGGC CACTTTGGGC ATCTCGGCGT GGCCTGCGCC GGGGTCACGG  
 2901 GGAGGGCTGT CAGCGCCAGG GCGGCGGAAC CCGAGGTCTC CAGACGAGTG  
 2951 AGGGAGGGAT GCAGGCTTGG GGGTGATGGA GCGCTTGGCT GGTGGCTGGT  
 3001 GAGCGTCCAT ACATCATAGC TCTCCTTCCC ACTCCCCGC CCCTCTTCGG  
 3051 GATTCTCTCT TTCTCTTTCC CCGTCCTCAT TTCTTTCTTC CTTTACTCAC  
 3101 CACTCGCTTC ATTCTCTTCC TTCCATTTCC TCTTTTTTTC TCCCCTCATT  
 3151 TCCTTTTTTT CCTTTCCTT TAAAGAAAG GGGAATCGTT TGTAACCCTT  
 3201 TCGTTCTACC AACGTGGAAT AGCTGTGAAA CCTGCAGCGT GGTCACTCA  
 3251 GCCTGGTCGT TTTCAGACCC GTCCTCATCC ATCAACATAT TTGTTTCCCG  
 3301 AGTCTATTGA TCTCCCTGAA TTCTACAGAA ATGCATTCTA AGCTAGGCGC  
 3351 CTGTATGTCA GAATCAGTTC TGCAGGTAGC TTCCGTGCTC CAAGTATGAC  
 3401 ATGTATTGTA AGGGCTGCAT CTGTTTTAAA CCCACATAAG CCATGGGTAT  
 3451 AAATAAATGT AGCTTTGAAA AAAAACTGG CTTATTCTA GATAAACTTC

Fig. 3 (cont'd 1)

53/ 124

3501 CCTCTTAAAT TACTGATATA CTCTTCTCCC TCTTTGACAT TTAATTTTAG  
3551 GAAAGTTGGG AGACAGGTTC TTGTCCTCCA GTTTTAAAGG AGCAGGCAAC  
3601 TTCTATTATC TTAATTTTCT CGTCTTTGAA CATCACTCAC GTTTGCAC TA  
3651 CCCAGTCAGT GGAACGAGTG GGTCATAATT AA

Fig. 3 (cont'd 2)

54/124

## Human genomic sequence

```
1  CCTGCATTAT TGTTTTTATC TGACTTCCAA TTTTGGTGTT CCCTGGGTGG
51  GTGGGTTTTTC CTGACACATT TACAAGATGC TTTTGGCAGG TTGGCTGGAA
101 TTTGAAGGCA CATTTAATTG TAGGTGCAAT AAAATATTCA TTTTCTCTTG
151 TTCTTGGTTT GAGATGTCAT GCCCTTTTGG TCACCTATAT TTTGGTGTGA
201 CTGTGTGTGT GTGTGTATGT GTTGTGTGA AGGATTTAAC AAAGTCTGTT
251 CTAAGTGTCA TGTGATTGA AGTTAAAAGG TATGTTAGTG ACAAGCCACA
301 AATTTCTCTT ATTTATAGTA CATTGATCCT GAAACCATTT TTTCCCTTGT
351 GATTTCCTCT GTGCATGGAT CATTTAACGA AAGGTTGGCA ATGATGAGCT
401 ATTTTCTTAT AATAGGAAAA AAATTCCTCA AGTTTACTTA CCAAGTCATA
451 TTTTATACAA GAGGGATTAG CAAATATTTT TGATCTAATA TTTTAATAGA
501 CTGAATTGCT GACCACTGCT AATTACCAAG AATATATTTT CTTAATTCTG
551 AAATTGCTGT ACCTCTCAAG TTGTCTGGAG GACTCCAAGT GACCCAACTT
601 GTAAGTCATG GCAACAGGAA GTGGTTGTTC TGGGTGCAAG CTGAAGTGTG
651 CACATGGACC CGTACTTTGT TAGCACTCGG GGACTTGATA TGGAAAGAAT
701 TAATGTACTG GCTTTTTTGT ATAGATGAAT GTTAACCTTC TGACATTAGT
751 CAGAACTACA TCTCCCAAGC CTTGTTTTGC AGTGTCTGTC CCTTTGCTCT
801 TCACTTACAG TAAGTCCTTA CTTAACTGAC TTGATAGGTT CTTGGAAACT
851 GCAACTTTAA GCAAAAGGAA GTATAATGAA ACACTTTAT CACAGGCTAA
901 TTGGTAGAAA CAAGACTTAA GTTCCCATGG CATATTTCTG GTCACAAAAA
951 CATTTCACAA CTTCTCAAAA CACTTCAATA TTAAGCATTC AAATACATGT
1001 AAAGTATGTA TATATGTAAG AAAGGTTACT ATAAACCAGA TCAATATTTA
1051 CCCAATTATT TAAGTTCAGG GTCTTAGGTG GCTGGAGCCT ATCCGAGTAG
1101 CTCAGGGCAC AAGGCGGGAA CCAGCCCTAG ACAGGACACC ATCCTGTTGC
1151 AGGGCACGTT CACACATGCC CACACGAGG CTGGGACCAT TTACATGTGC
1201 CAATTACCTT ACCATGCACA TCTTTGAGAC GTGGCAGGAA GCAAGAGTAC
1251 CTGGAGAAAA TCCATACAGA TATGGGGAGA ATGTACAAAC TCCACCCAGA
1301 CAGTGGACCC AGCCAGGAAT CAACATTTGG GCAACATTAT AATGAAACGA
1351 AGTTGAATGA AATGATGTCG TTCCACGACC TGCTGTACTT GAGGGGTGTT
1401 ATAAAATTCT CAGAAGACAG AGGTTTAATG CTATCTTTT AATAGAAAAT
1451 AACTTATAGA GAAGTGTGCA CATGTGACTT TGTGTGTAGC AGGAATCATT
1501 AGGATGAGAA TCAGACGTAA GAGGTGGTGC CAACATGAGG AATGTTGAGA
1551 TTCAGGGAGC TGTGGATGGA AGTAGAAGCC AGAAGGCCAG GGTTAGGTTC
1601 CTACTTCTTA CTGTTTCAGT TATTGCAGTG TTGGCCTGTT TATTACAGA
1651 TGTACCTTAG CTTTGTTTTC TCAAGAAGAA AAATGAGCAT AATCTTTCCT
1701 GTTATGAATT CTTAAACACA CAGGACATAA CCACAGACAC AGAGGTGCAC
```

Fig. 4



55/ 124

1751 ATATGTAGCA GTAATGGATA CTAAATGATA CACTCGGAGG AAACAGAAAA  
1801 GACTTCTGAA TAGAGACTGG AGATACTTCC TTGGACCATT GATGAATGGG  
1851 CAATGATGCA TTTTGTCTT CCATTCAGAA GGCTAATATA TTGCTCTCTA  
1901 TGTTCATATGG ATAAAGGCAG TATATGCTCA AGGATGAATC ACATAATATG  
1951 CATAATAAAT CCAGCAAGCA TTACCCCTTT ACTTATGTGA CTGCAAGTAG  
2001 GAATACATTT CCCCCTCTT AACCATGTAA GATTTCTTTC CCTTCTCCCA  
2051 TTTTGTAAAG AAAAGTAAGT TCCTGAAAGG TTAAATGGAC CTCAGGATGG  
2101 GAAAAATCCC CAGAGCTATC TTTCTGCACA GACTTCATTT TTTCTCCCAA  
2151 GTCTGACTGT CAACTGCGAT ATCTGATATG AGGCTCTGGT GCTGATGTTT  
2201 CCATAGGTCA TCATCCTTCG GTGTCCCAGA TGAAGTCTCA GGTCAACAT  
2251 TGCAATAGCA CAGATTCTGA ATTTAATGCA TCATTAAAGT TGGTTATGTA  
2301 ACCCAATGGC CTTGTAAAC TCCAGATTTT TAAATTTATA TGTATTTACT  
2351 ATTCTCTTAT TTTAGAAATGA TCTCACAATG TTCACAAGAA ATAAGCCAG  
2401 TCCCTGCAAA GACTTTAAAA GCTGCTTGTT CACATCATTA GATTGTACAA  
2451 CGCTTGACAT ATGACACTTT TTGCTAATCT ATGCAACATT TTTGTAACAA  
2501 TTGTGCACAT TTTAACTACT TCAGATAATC AGGACCTAGA GACTTCAAGA  
2551 TCTGGAAGCA TTGCTGGTGA CATAGAGCAA AAATTTCTT GAGAATAGGA  
2601 AGTCAGTGTT TTGACAAGTG ATTTATAACA GTTCAGGTAT AGCCAGGAAG  
2651 GTTTGAAACA AACCTTAAGT ATTATTTCTT TCATCTTGAT TAGTATATAT  
2701 TTATATGTGA TCTATTTATG TATATTAATA GATTTTGGG TCTTATAGCC  
2751 AGCTTTCATT TTTCTCTATT GGAAAAGATC TAAGTCCCA TCCTTCCTTG  
2801 GTGGCTTTTG GTAGGTTTGT AGACAAAACA TTGAAGAATC AATGGTACCT  
2851 TTTATACATT AATACTGCCA ATATGACCAT AAAATCATAT TTTTGGGAA  
2901 TTTATTCCCC CGATCAAAAG AAGCATTTGT TATTGAACAC AGTCTTATGC  
2951 TACCTTATTA AGATGTATCA AACACCCTGA TTGATCAAAA ACACCTCAGT  
3001 CCATTTTAAG GCAGTATGTC CCAGCAATTA AAGATGTAGC TTCTGGAGGA  
3051 GTCTTTCTGA GTTTGAATTC AGTACTCTC CACGTACTAT ATAGGTGATC  
3101 TTGGGTAAAC TTCTTGAGTC TCAGTATCCC CATCTGTAAA ATTGTTGTAG  
3151 AGAAGAATTT TTGTGATGAT TAGGTGAGAG AATATATTAA TGTAATATTT  
3201 AGGAGAGCAA CCAGCATGTA GCATATATTC ATTACATATC AATTTCTATA  
3251 TTATTGATGT TCATACTGCT GATGTTGAAA TGCACAGGAA GGCCACAGTT  
3301 ATTTTCTGTT TAGATTGATT TTTCTTTTAA AGTCTGAACA TAAACTGTAA  
3351 TACTGTGCTT ATTTATGTAG GAACTGTGAT CTCGTCTCCT CCTTTTCCCA  
3401 TCTCCCCCTC TCTACCTTAG TTTTCTCTTA TAGTCTCAAG CTGAAAACAA  
3451 TGACCAGGTG CCTAAGAGAT AAGAATACTC TTTCTTTTGA ACTCATGGCA

Fig. 4 (cont'd 1)

56/124

3501 TTAGCAGTGA CCTGGATGAG ATTGGAGGCT ATTATTCTAA GTGAAATAGC  
3551 TCAGGAATGG AAAACCAAGC ATTGTATGTT CTTACTTATA AGTGGGAGCT  
3601 AAGCTATGAG GATACAAAGG CATAAGAATG ACACAACAGA CTTTGGAGAC  
3651 TTGGGGAAAAG GGTGGGAAGG GGGTGAGGGA TAAAAGACTA CAAATAGGGT  
3701 GCAGTGATATA CTGCTTGGGT GGTGGGTGCA CCAAAATCTC ACAAATCACC  
3751 ACCAAAGAAC TTACTCATGT AACCAAACAC CACCTGTTCC CCAGTAACCT  
3801 ATGGATATAA AAAAATTAAA AAAAAGAAAA AAAGAAAAC TTTTTTGCA  
3851 GGGGGCAGGT AAAGGGTAAG AGGGCATCCC ATTTTGTAGT TTCTAGAAAA  
3901 GCTT

Fig. 4 (cont'd 2)

57/ 124

## Human genomic sequence

```

1  CTGCAGGAAG CAGCAGCAAG GTCCAGGGAG CCTCTAATTT AAATAGGAGA
51  AGTCAGAGCT TTAACAGCAT TGACAAAAC AAGCCTCCAA ATTATGCAAA
101 TGGAAACGAA AAAGGTAAGT GTTTGTTACA TCATTATGAC ACAAGTCCAA
151 CATGAGTCTT GTGAATTGCA TGCTAAATCT AATATTTGAG CAGCGTAACA
201 ACTTTGGGCC TAGAGATGTT ATCAGTGGAG TTTCTTTATG TTTCTTAECT
251 GTCCCCCTCT GACTGCCAGC TTTCTTATCT GAAGAACATT TTAAACAAAT
301 AAACCTCATC ATTTTAAAGT AGTTAGTTAT ATATGCAAGT ACAAATACTG
351 TTTCTCAAAA ACAGGTCCTT CCAAATGCAT GTAAATCACA TTTTCTTATG
401 TCTTTTTATG TTTTTGAAAA TGTATCCTGA AATCATAAAG CCATATTGAA
451 TTTATCTGAA TCCTTAACTT CAGTTAAGGT AAGAGCCATA AGTGTTTTTG
501 ACAATTAAGG TTGGAGCATC AAAATTTGAA ACATAATTAC AGTAGGTTTT
551 TATCTTTGCA AGCAGCAGAT CCCAGAGATA TTATGACCTC AGTTTTCCTC
601 AAAAGACAAA TTATTCATAT TTGTTTTGTT TTCTTGAATT AGTGCATAAT
651 ATAAATATCA AATCACAAA TCAAGGACAT TAAATGAAAG TGTCTGTTAA
701 AGGCATATTA TAAATGAATC ATAAGCCACA CAGTTCTCTG TGATGTACGA
751 AGTGGGCATT TAAAGAGGTG CTGATTTGAT GCTTGTCACT GAGTAGCAGA
801 GAGGACGGGG ATGAGTATGT GTAGTTTACA CCTCAATCAT GAGGAAGTGA
851 AGAACTTGTT CTGTTATAAG TAGTATGGCT GTGTGAGGAA CTAGGGTGTT
901 CTGCTGGATT TTGAGGAAGT ATTTTCAAAT CAATAGAACT TCAAACCTTT
951 CTTCAGAGTG TTGGGCTCTA CATGGAAAA CACATGAAAT TAAAAAGTGG
1001 CACAAATGTT TAGTTAGTAG AACATCTGGC TAATTGGGAT CAAATAATTC
1051 AACCATGTGG GAACGTTTTT GCTCAAAATA GATAATTGTG AATTGTTTCA
1101 TATAGGCAAA TGATTAGACA ACTTCCTCTT CCTCAAATGT GAACGGACAG
1151 ATGTGATCTA GAAGCAAGAC ACTCTTTTGT GTAAATATTC CCTTTGGCCT
1201 AAAGCAAAAG TGGACAGACT TTAAACACCT GAGAGCAGAG CAGTGTGTGT
1251 TAAGATTGCA ATATCTTAAG CTCTTGAGTT AAATGGAAAA TGAAAAACAA
1301 AAGTGTATAT TTGGAAGTTA GGAATGTTTT CTTTAAAATA TAAAATAAAA
1351 TTTTAGATTT AAGATCACAA GAAATATTAC TGAAGACTTA TACTCTTCCT
1401 GGGGCTAAGG GAGGTGACAG TCGCTCATCA GAAAAAAA AATGCCCTCA
1451 TTTCTTAECT TTTCTAAAA ATATAATACA AGTTCAGGCT AATACTTCCT
1501 GTATATGTGG GAAATTTCTA GGGGAAGCTA ACAGGCTTAG AAATAAAGAT
1551 GTGTTAAATA GACTACCAAA GTGTCCAATT AAGCAACACG ATACCACCGT
1601 TATTGATATT CTAGCAAGAA ATTACTAGCA ATGTTTGTA AATAGACTTAG
1651 AAATGCATTT GATGAATTAA CACTTTTATA TCTTAATTTA TCTGAATTTT
1701 TCTGTAATGT GAAAATGTTT TATTAACTT ATTTCTGGCA TCTATTAGTA

```

Fig. 5

58/124

1751 AAATTCTGAT GATATACAAG CATTAATATT TTTCCATGGC CACTCAATTC  
1801 ATACATACCT TCCCTATCTA TGCTTAGAAG GCAGTGCAAA ATTAGATAGT  
1851 AGCAATATTG ATTATAACCA CAAGGTGGAG ACAGATGTCA TGTAATATGC  
1901 AGTCTGCTCA TATAAAGCAC ATTTTCCTAG ACAAGAGTTT TCATACGATA  
1951 TAATAAAGAC ATCTGGAATT TGTCTTGAT GCAATATGAA ATTTGCTATT  
2001 AAACGTGGAG TTAAACTTT ATGTCAATAG ATCCAATAAC AATGTTTCATA  
2051 AATTAATCAT TATGTCATGC TGTATTTCCT AAATACTATC TTAAATTATA  
2101 AGAGCAAACG AGGTAATAA

Fig. 5 (cont'd)

59/ 124

## Human genomic sequence

```

1  GTACATTTT TAATAAGAT GTTGTTTTA ACTTTTGA TATGAAGATT
51  TCTAGTTCTA GAATAATGTT TATAAAAATA TACAAATCCA TCTGGTGATG
101 AGTTGACCTC TATCACAAC TTTTGCATA TATAACTTGG GTGTGACCAA
151 GCAAGGTGAG AGTTAAGAAC TTTTAAACT TACTGTATTA TATTGATAGA
201 ACTCAGAAAG TACTAACTTG AATATTATTA TTCTAATTGC TTTTCCCTTT
251 TAGTTATTAA AAATAAGAAT ACTTAAATTA ATAACAAGAT CTTTACTGG
301 CAGGATTAA CAAATTATCT GTAATGTGTT CCTCGAATGC TTTTAAGTGG
351 AAATATACTT TATACATTCT TTAACAAC TCAGAGGATG AGTTACATAA
401 ATCAGTTCAG GAATCTATAG AATCTGTAAT ACATAGTAAA GGTTTATTCA
451 CAATTAAAC AATTTCACTT CTATATTAA AAAACAAATT GTTGAAAGTA
501 CAGTGGCTTT TCATATGTAT GATTTGTAAA ACAAATTAGC TTTTTTAAAG
551 TGATGTGACG CTTAATGAGA AGAAATCAGT AGAGAATTAC AACTGCACT
601 TCAAAAGATA CATCTAATAT CATTTTAATA ATGAAATTTG AAAAAATAGT
651 GTGCTCGTTT TACAGTCTCA TTAAATGAAT TAAATATCA GCACACATTG
701 TAGTAGGTTA TCATTGGCAG AGAAGGCTGA AATAGAAACG TTACAATGGG
751 ATGCACTGCC ATCTGAACAT TATGTGCAAG TGGAACGCGG AAACATATTT
801 CTCAGAACAA GTGGTAAAT GAAAACAGCA TCATTTGTAA AGCATTTCCT
851 TTGAGAGTGC TTCAGTTTCT TCTCCTGATG ACCTGCCATT CAGAACTGA
901 CAATGAATAA TACACTCTGA CACCAGCATT TGTCAATTTG CCCAGAACCA
951 TATGAGAGTA CTCTAGACAG ATATATGTTT CGAAGTAAAC CGAATACCTG
1001 TTAAGTGTAA ATCAAATCTT GTAGAAACCA TGCCATGGTT CCTTTGGACA
1051 TATACTTTGC ATGCCTGAAG CAAGTTACCT TAAGAAATCA TTCTTTTGTT
1101 TTACAAAAC TGTATTAAAA AATTAAAAAT GCAAAAAGC TTAATATTAT
1151 TAGGAATTTA TCCATAGCTT TATTTGGAAT CCAGTTTCTT TATTATGATC
1201 TATAAACATG CATCATTTGA TGGAGTTCCT TAGTGGAGAG GTGT'TTTTCC
1251 ATGTTGTCAA GAAACATGCC CCAGCACCAG AAGGGATACT ACCTACCATC
1301 TTTTGGCCAT TTCTCACCGT GATTCTTACA TTGTACCTGT TTACTCACTG
1351 AACAGGGCTT CCTTCTCTTT GTCTAGATTC TAATCAGGTG TCTTCTGGTG
1401 TGGAAGCTTT GGCTTTTATT TACACACAAC ACAGAATTAA TAAGATAGAT
1451 GCCAAGGATT TAGCAACATT TTAATTCAAC ATTATACAGG TATCAGAGTT
1501 AATGAGAATT ATGCATTAGT CTTTAAATTT GGGCAGCTTA TTCAGCTAAA
1551 ACATAGATGT CTAGCTCTTA AACACTTTGT TTTTTTAATT ACTCTGAAAT
1601 TACAATAAAG TCAAAGAACT GAACTGTTTT CTTTCAAGC CAGTGCAAAT
1651 GTGCTTAGT TATTATTTTA CTGGTGATCT AATTATGCAT TTTAATGCTT

```

Fig. 6

60/124

1701 TATTACTTAA TACTTATATA AGCCTAAAAT ACGTTGTTAA TGTCATAATT  
1751 TCAGGGATTT TAGTATTCTT TCCATGAGTT ACCATAACTA GGTGCATATG  
1801 TGTAAATATA CGTATATATC TATATCTATA TATTTATATC TATGTATATA  
1851 TCAATTTATA AGACTAAATA GACTTGGCCA TATGTGTTGT TGGTTTATGC  
1901 ATACATGCAC AAATATTGAG GTGTCCACAA AGTATATATG CCTGTACATA  
1951 AATTACATAC TGGCTGGTGA GTGAATGTAA GCTTCTCTAA ATTGTACAAC  
2001 TCTCCACAGA GTGGCACTCT AATATTGCAA AGGTACAATA TAAGCATGTG  
2051 CAGAAATGAAC AGCTCTTCTA GGATCCCTAT AAAACTCCAC CCCATGTTTC  
2101 TGT

Fig. 6 (cont'd)

61/124

## Human genomic sequence

```

1  AAGCTTCATC CCAGAGGGGC ACTTGCCAGA TGCCTGCTAG AGCTCTCCTG
51  TATGAGGAGT CTATCAACAC CTGCTGGGAG GTGTCTCCTC GTCAGGAGGC
101 ACGGGGGTCA GGGACCCACT TGAGGAGGCT GTCTGTCCCT TAGCGGAGCT
151 AGAACACTGT GCTCGGAGAT CCGCTGCTCT CTTCAGAGCT GGCAGGCAAG
201 AGTGTTTTAG TCTGCTGAGC CTGCGCCAC AGCCGCCCT TCCCCAGGT
251 GCTCTGTCCC AGGAGATGA GAGTTTTATC TGTAAGCCCC TGACTGGGGC
301 TGCTACCTTT CTTTCAGATA TGCCCCGCC AGAGAGGAGG AATCTAGAGA
351 GGCAGTCTGG CTACAGCAGC TTTGCCAAGC TGCAGTGGGC TCTGCCCAGT
401 CCAAAATTCC CAGCGGGTTT GTTTACATTG TGAGGGGAAA AGCACCTACT
451 CAAGCCTCAG TTATGGCAGT TGCCCCCTCC CCCACCAAGC TCCAGGTCC
501 CAGGTGTCCT TCAGACTGCT GTGCTGGCAA TGAGAATTTT AAGCCAGTGG
551 ATCTTAGCTT GCTGGGCTCC ACAGGGGTGG GATCCACTGA GCTAGACCAC
601 TTAGCTCCCT GGCTTCAGCC CCCTTTCCAG GTGAGTGGAT GGTTCTGTCT
651 CACTGGCATT CCAGGTGCTA CTGGGGTATG AAAAAAAAAA CTCCTGCAGC
701 TAGCTTGGTG TCTGCCCAGT TTTGTGCTTG AAACCTCAGC CCTTGGTGGT
751 GTGGACACCC AATGGAATCT CCTGGTGTGC ATGTTGTGAA GACTGTGGGA
801 AAAGCATAGT ATCTGGGCTG GATAGCTCCG TCCTTCAAGG CACAGTCCCT
851 CATGACTTCC CTTGGCTAGG GGAGGGAGTT CCCCACCCT TTGCACTTCC
901 CAGGTGAGGC AACACCCAC CCTGCTTCTG CTCACCCTCT GTGGGCTGCA
951 CCCACTGTCT AATCAGTCAC TGTGAGATGA GCCTGGTACC TCAGTTGGAA
1001 ATGCAGAAAT CACCTGCCTT CTGTGTTGAT CTCACTGGA GCAGCAGACT
1051 GGAGCTGTTC CTATTCAGCC ATCTTTCTCA GGTCATAATC ATAGATTTTT
1101 AATTGATCCC AGCAACATGG ATTAGTAAAC AGCATATTTT CAAGTGATTT
1151 TTTTTTATTT TAAGGTCAAA TCTACAAAAT ATTATAGTGT TATCACCCT
1201 TAAATTTATT ACTGGTGATA CTATGTTTGT CTCTATTCAC ATTTTATTGC
1251 TAGAAAGAAT TATAATTTGT AGATAATAAT AGTTATTTGA AATGTATTAC
1301 ATATCCTTTT ACTTTTAAGA AGAGGTGACT TAATTATCTA GGTATACAAT
1351 TATTTTGAGG ATACTAAATG TCATGAATAG CAAATTTATC ATATTGCTTT
1401 CCTAGGTGAA GACCCTGAAA CAAGAAGAAT GAGAACAGTT AAAACATAG
1451 CAGACTTGAG GCAGAATTTA GAAGAGACTA TGTCCAGTCT TCGTGGGACT
1501 CAGATAAGCC ACAGGTTTTT TTCAATTTTG CATATATTTG AGCCAATAAA
1551 GAAAAAATAA TTACAAACAA ACATTTAACT TTTCTTATAA TGACAGAGAT
1601 GGGATTTTCA TTTCCCTTA CTATTTTCTC CCTTGTTTTA TATCAAATTG
1651 ATTGGTAATT ATCCTTAAAC TGAGAATTCA CAGTATATAC CTATTTATCT
1701 TTTATCTCTA TCTCTATCTG CTATTTATGT CTTTTTCAGT ATAATTTCCA

```

Fig. 7

62/124

1751 GTACTGCAAC TACCACCATC ACTGTTAAGT GGATTTGTAA TACCTGTCCT  
1801 AGAAAACAGT GGCACAAGTT GCACTTGAAA TGCATCTGGG CAGGGTAGTA  
1851 GGGAGACATT CAAACATAAT TGTAAGTTAA TTTCAGAATA GGTCTGGGAA  
1901 GGTACAGTGT AGTTAAGGAT TTGTTGAAAA TGTAACAA TATGTTGTTT  
1951 TACCCAAGGT GTACTGATGG CCTTTCTTTT GAAAACAAAC GAAAAGCTAT  
2001 AAAATGTATG CCCCTTTCCA CAATTTGACC TCAAAATGAA TATAGAGTTT  
2051 AGCTTTCGGG AAGATGACGT GTTTATAAGA GATGACCCCTC AACTCCAGCC  
2101 TTTTCTGTCT TCATGCATTC TAGATTATGG CCCTAAGTGA ACCAGAGTAT  
2151 AGTTATTTCT CCATTTTATT TGACAGCACC CTGGAGACAA CATTTGACAG  
2201 CACTGTGACA ACAGAAGTTA ATGGAAGGAC CATACCCAAC TTGACAAGTC  
2251 GACCCACCCC CATGACCTGG AGGTTGGGCC AGGCATGTCC GCGACTTCAG  
2301 GCGGGAGATG CTCCCTCCCT GGGTGCTGGC TATCCTCGCA GTGGTACCAG  
2351 TCGATTCATC CACACAGACC CCTCGAGGTT CATGTATACC ACGCCTCTCC  
2401 GTCGAGCTGC TGTCTCTAGG CTGGGAAACA TGTCACAGAT TGACATGAGT  
2451 GAGAAAGCAA GCAGTGACCT GGACATGTCT TCTGAGGTCG ATGTGGGTGG  
2501 ATATATGAGT GATGGTGATA TCCTTGGGAA AAGTCTCAGG ACTGATGACA  
2551 TCAACAGTGG GTAAGTAACC CTGTTCTCCG TCAGCATTGT GTGAAGAGGG  
2601 GAGGTGGTCT ACTATAATGC ATTCACTATA AACAAATGTG TAAGTTTGCC  
2651 CAGAAAGTCA TGAGAACATA TGAGATATCT GAGGTTATTC AGAGTGTTGA  
2701 AGGGCCCTTC CTCTGCTCAT TCATGGAGAG TAAAGAATCC AAGATTTCTA  
2751 TAAATTCATT ATAAGCCGCT AAGTTTTTCT GTTGTGAGA GAAACACATG  
2801 TGGCTTCTGT TTTTCAGAGT GATTTTCACA TGCTTCTTAA GTAACAGATT  
2851 TTGTAGTTAA GGACGTGGGA AGGAGACAGG AGGAGTTTGT CTGATTTGCT  
2901 TGATTTTTTT TTTCTTTTTT AGCTTGTTAG AAGCGGCCTG TAAGTGCTTT  
2951 GAGAAACAAA TATTTTCTTA CTGCTTCAA TTATGCATCC CCAATTTAAC  
3001 TTGAGGGAAA AATCACTTTG GAGTTGAAAG TTCACTCTA TTCATTTTCT  
3051 TTTGATGGTA TCAGATTTCA ATACATCTCA GACCTGTTT TTCTTCTGTG  
3101 TCCTATTACA TTCCAAAACA TGTGTGATT GTAAACTCT TAGAGTATAT  
3151 TAACAATTTG GGATATTTGG CATAATCAGA GAATAGGTCC AAAAGGAGGC  
3201 AATAGGATAT TCTATTAATA ATTGTAATTG CCATTTTGTAG CATTTCTGTG  
3251 TATGTACTAT GCTCTTGTC AAGTCTTTGA AGATAGTGT TTACTTTTCC  
3301 TTCCACCAC CAGCAATGTT TATGAGGTAG ATGTTTTTAT ACATGTTCTA  
3351 TGGATAAGGA AACTGAGTCT AATTGGCCCC GGCTGGGAAC TAACGCTAGG  
3401 GAAACGGCAG ACCTGCATTA GAACTCAGCT ATGTCTGACT TCAAACACAG  
3451 GCTCAGTAAT ATGTGGAAAA GCTTCCCAAT TAACTTTGTC TATAAACTTT



63/124

3501 GTGTGAGTCT GGATTTTGAC TTACTCTTTG TCTTTACGCA TCTGAGAGGA  
3551 CCCATGTAGG AAATAATTCT TCTATATAAG TGACCCTTCC TGACTTCATT  
3601 CATGAAAAGC TTATGTTTGA AGGGTGACAC GACCTAAAAA AGAGTACAAA  
3651 ATAGCTTTTG ATTACATTTA TAGCTTTGCT CTGATATCCT AATACCTACT  
3701 AGTCCATTCC TGGTATCCAC CCTACCTGAC TTTCTAAAAA TTTAGAATTA  
3751 TAGAGACTAA TTATGATTAA TTAAGATAGG TTGTTGTTCA GTTGCCACTG  
3801 GATTGAGAGT GCCTAGTTTG AATCTCTCCC ATTCACTATC TGTGGACCCC  
3851 TTCGGAACCT AACGTATCCA AATTAGTTTT TGTCATCTAG AATAAGGATA  
3901 AAATTGTACC ATCTTCATGA AGTTGTTAGG ATCATCCACA AATTTTAGTT  
3951 TGCGCAATGC TTGGCATGAT ACAAGCACTC AATAAATTTA TCATCTTCCT  
4001 CTTTATCATC ACTATTACAT TTATTATCAT TAATAACCAT ACCAATTTTT  
4051 GGTGTTGTT AGTTATAATT ATCATTTTTG TATGTATTTA ACATAGCCTA  
4101 GGAGGCAATG CCCAGTTCAG AAAACATAAT GGCAAAGCAA GAGTGTCTAA  
4151 GGCACACTCT TTCTCCCATC TCTCTCTTCT TTCTTCTCCA TTCTTTCCAC  
4201 TCTATCCCTT CTTCTCTTTT TTTTCTCAAT CTCCTTAGAT GTGGACATAT  
4251 GTGTGAATTC

Fig. 7 (cont'd 2)

64/124

## Human genomic sequence

```

1  TGTGGGTGTG GGTGTGAAGC ATGTGTATGT GTGTGTGTGA AGCATCTCCC
51  CACCTGTAAT GTAAGTCCAT GAGTGCAGAA TTTTGTGACAT ATTCTTTTACG
101 TGTTGAGTTT TAACAAATGT TTGTGGAGTG AATGAACAAA TTAATGAATA
151 TAGGCTATTT ATTAATTAGG CAATATAGTC ACATAGGCTG GCAATCGCAT
201 CTAATTAAAT AGAGTGGTAA ATGAGTTCCA GAAAGAATA AGGTACTACA
251 AGGATGTTAT GAAAGAGAAA AATGAGTTAT GTGAAAAATA GGAGACAGTG
301 ATAAGAGGGA AAGAATCCCA AAGTGTGGGC CACATTTTGA AACTAATGAC
351 CTATTATTCT ATTATTGTTA GCTGAAAGTA GAAAACGTCA TGGGAGGGAA
401 TATCTGCTAG TTTTGGTAA AGGATGTTGT GATGGCAGAA CCAAGAAATG
451 AACACAAGGT GACTTTGGTT TGGGGACAGT GGGATAATCA ACTCTCCTTG
501 CTCCATCAGG GCCCCAGACT GGGCTCTGGC AGAGGAATC AGAACAACGT
551 AAAGACCTAG ATAGGTATCT AATAAATTGG GACCTGTGAA AACAGTGCCT
601 CTTAAAGTGT GGTACCTGGA CCAGCAGCAG CAGCAGCAGC AGCCATTGAA
651 ACTTCATAGA AAGACAGATT CTCAGCTTCA TCCAAGACTT ACTGAATTAG
701 AATATCTCAA GGTAAGGCCT GGTAATCTGA GCTTTAACTA GCCCTCAAGG
751 TGATTCTTAA GTTCAAGCAT CACTATATTA AGTTGAACAA ATAGATGCCA
801 GGCCTATAAA TACATGTAAC GCCTAGCATA AATATTTCAA CATTAAAAAT
851 GACATTTTCAT AGTTCTTATT TACCCTATTA GCTGTGTTCT GTCAAGATAA
901 TGAGAATATT GATATGTTAG AATACACTGA TGCACATAAT TTAAATTAG
951 ATCAAATAAT GACTTGTTAT ACCTGAAATA AATTGGTTCA GCTTGGTAGA
1001 TGCAGTTTTT GAGAATTATA TAAGTCATTT TAAAAGAAT AATTTTAACT
1051 TGAGCTGCTT GCATAAATTA AATTGCAAAA AGGTCATAGT ATAAATCCTC
1101 CTATTAGCAG AGATAGAAGG TTTTAAAAA AATTACAGAT AAGTCTGAAG
1151 GTCTTTTAAA ATCTTATATT CAGGAAGTGA CTCGGGATGT ATATCATTTT
1201 AAAATACATG GTCTTAAATG TTGTAGTTGT ATGACTCTTT CAGTTAATTT
1251 AAAATACTTC CTTCTATGAA AAATTGTTTC AAAAATTTT CTAAATTCGT
1301 TTATCCATTT CAAGTAGGAT AGGCAAGAAC AGATATAAGA TACTACTTTT
1351 TTGTTTCATGT TTACTAAAAA AAAAATTACT GTAATTGAGA TCATGTAAAA
1401 ACATGTTTCC TGTCTATTTG TCTTAACCTT TTAATCCTGG CACCTTAAAT
1451 TTGACATAGT AGGAATTAGA AGACAATTGC AGAAAATGTC AACTGGGGAA
1501 ATTTTATTCT ACTAAAACT ATGTCCATAC AACATAGCAA ATCACATTTT
1551 AAAGGCCAAA AAGTCTTTCA TAGCAATTTT TCAGATTATT TTCAAAGCAT
1601 ATCTTCTCTC TGCTCCTGCA GCATGCCGTT GATTTTCTG TTATGCAGTC
1651 ACATAAGTAA TTACATGTTT ACATGTCTAT TTCCTCATA GAACACGAAA
1701 CAGTTAAATG TAGAATAATA TCCAATCCAT CTTTTTATCA CCAGTAGCTA

```

Fig. 8

65/124

1751 GCATACTGTA GGAAGTCAAT AAATATATCA GATAAATTGT GGAAATAACC  
1801 ATATCAGCTT ATAACATATA GAAATGTGAG TTAAAAAGA AAACAATTAT  
1851 ACATATGAAA AAATTTTAT ACCATTTTTT TAAAGACCTT TCAGATGTCA  
1901 TACAGTTTGG ACTTTTCCAG TGTTTCTTGT ATCATGAGAC AATAGTAGAC  
1951 ATTGTAAATC AAAAATAGTT TTCTGGGGTT GTGTACATTT GAAAAAACTG  
2001 AATATCATAT CTGTTCTTAG AGAGTAATGA TGGATATTAA CATATCAAAG  
2051 GTACAGAGAA GTCTTAAAGT TCAAAGTAAC ATCTGCTTAA TTGTATTTAA  
2101 TTCAGTGCTC CATGAGCTTT TTTATCACTG ATTCCCTCCC TTTTCTCTCT  
2151 TATGATAATA ATTAAC TTGT TCCTGTAGCA TTTTAAGAAA TGTTGATTTA  
2201 GTTGAATGCC TTCACTTCTC CAATATAATA GCAGAACTC AGAAATATTT  
2251 ATTTACCCAG AATCATGCAG CTAATAGTAC AAGGATTCAG GTCTTTTACT  
2301 TCCTATTTTG TGGTTCCCAA CTACTTTTGC CAAAGGTCTT TTAAATAATA  
2351 TGAAACATAT TAGTGATTGA TTCATTATAG TAAATGGGTA AATGATAAGG  
2401 CTTGCAATAA TTCACTGACA AGAAAGCTT

Fig. 8 (cont'd)

66/124

## Murine cDNA sequence

```

1   AAGCCACAGCACCCCTGGAGACAACCTTTGATACGACTGTGACAACCTGAAGTGAATGGAAG
   S H S T L E T T F D T T V T T E V N G R
61  GGCCATCCCCAACCTGACAAGCCGACCTTCCCCCATGACCTGGAGACTGGGTCAAGCGTG
   A I P N L T S R P S P M T W R L G Q A C
121 CCCTCGTCTACAGGCTGGAGATGCCCCCTCCATGGGCGCTGGATATTCTCGAAGCGGTAC
   P R L Q A G D A P S M G A G Y S R S G T
181 CAGCCGATTTCATCCACACGGATCCCTCCAGGTTTATGTATACCACGCCTCTCCGCCGAGC
   S R F I H T D P S R F M Y T T P L R R A
241 TGCTGTCTCGCGTCTGGGAAACATGTCAAAATAGATATGAGCGAGAAAGCAAGCAGTGA
   A V S R L G N M S Q I D M S E K A S S D
301 CCTGGATGTGTCTTCTGAAGTGGATGTTGGTGGATACATGAGCGATGGTGATATCCTTGG
   L D V S S E V D V G G Y M S D G D I L G
361 GAAGAGTCTGAGAGCGGATGATATCAACAGTGGGTACATGACAGATGGTGGGCTCAACCT
   K S L R A D D I N S G Y M T D G G L N L
421 ATATACCAGAAGTCTTAACCGAGTCCCGGACACAGCAACTTCCAGAGATGTCATACAGAG
   Y T R S L N R V P D T A T S R D V I Q R
481 AGGCGTTACAGATGTGACAGTGGACGCAGACAGCTGGGATGACAGCAGTTCTGTGAGCAG
   G V H D V T V D A D S W D D S S S V S S
541 TGGCCTCAGTGACACACTTGATAACATTAGCACAGATGACCTCAACACCACGTCCTCCAT
   G L S D T L D N I S T D D L N T T S S I
601 CAGTTCTTACTCCAACATCACTGTCCCCTCCAGGAAGAACAACCTCAGCTGAAAACAGATGC
   S S Y S N I T V P S R K N T Q L K T D A
661 GGAGAAACGTTTCACAACAGATGAGACCTGGGATAGTCTTGAGGAGCTGAAGAAAGCCGA
   E K R S T T D E T W D S P E E L K K A E
721 GGGAGATTGTGACAGCCATGGTGACGGAGCCGCCAAGTGAAGGGTGCTACTTCTGGACT
   G D C D S H G D G A A K W K G A T S G L
781 TGCTGAAGACTCGGAGAAGACAGGGCAGAAAGCCAGCCTGTCTGTGTCACAGACAGGCTC
   A E D S E K T G Q K A S L S V S Q T G S
841 CTGGAGGAGAGGCATGTCTGCCAGGGAGGAACCTCCAGCTACAGCTAGGCAGAAAACCGAG
   W R R G M S A Q G G T P A T A R Q K T S
901 CACAAGTGCACTCAAGACCCCTGGGAAGACAGATGATGCCAAAGCTTCCGAGAAAGGGAA
   T S A L K T P G K T D D A K A S E K G K
961 AACTCCTCTCAAAGGATCATCCTTGCAAAGGTCTCCTTCAGATGCAGGGAAAAGCAGCGG
   T P L K G S S L Q R S P S D A G K S S G
1021 GGATGAAGGGAAAAAGCCACCGTCAGGCATTGGAAGATCGACAGCCAGCAGTTCTTTTGG
   D E G K K P P S G I G R S T A S S S F G
1081 ATACAAGAAGCCAAGTGGTGTAGGGGCTTCCACTATGATTACCAGCAGCGGTGCCACCAT
   Y K K P S G V G A S T M I T S S G A T I
1141 CACAAGCGGTTACGCTACACTGGGGAAAATCCCCAAATCCGCTGCCATTGGTGGGAAGTC
   T S G S A T L G K I P K S A A I G G K S
1201 CAATGCAGGAAGGAAAACCAGCCTGGACGGGTCCCAGAATCAAGATGATGTTGTCTGCA
   N A G R K T S L D G S Q N Q D D V V L H
1261 CGTGAGCTCGAAGACCACCTCCAGTACCGTAGTTTGCCCCGCCCTTCTAAGTCCAGCAC
   V S S K T T L Q Y R S L P R P S K S S T
1321 CAGCGGAATCCCTGGGAGAGGTGGCCACAGGTCGAGCACCAGCAGCATTGATTCCAATGT
   S G I P G R G G H R S S T S S I D S N V

```

Fig. 9

67/ 124

1381 CAGCAGCAAGTCAGCTGGGGCCACCACCTCCAAACTGAGAGAACCGACTAAGATCGGCTC  
S S K S A G A T T S K L R E P T K I G S

1441 AGGGCGCTCGAGTCCAGTCACTGTCAACCAACAGACAAAGAGAAGGAGAAAGTAGCAGT  
G R S S P V T V N Q T D K E K E K V A V

1501 GTCAGATTGAGAGAGCGTTTCCTTGTGAGGTTCCCCCAAATCCAGCCCCACCTCTGCCAG  
S D S E S V S L S G S P K S S P T S A S

1561 TGCCTGTGGGACTCAAGGGCTCAGACAGCCAGGGTCCAAATATCCAGATATTGCCTCGCC  
A C G T Q G L R Q P G S K Y P D I A S P

1621 CACATTTTGAAGGTTGTTCCGGTGCCAAGGCAGGCGGCAAAATCTGCCTCCGCACCTAATAC  
T F R R L F G A K A G G K S A S A P N T

1681 TGAGGGGGCGAAGTCCTCCTCAGTAGTGCTCAGCCCTAGTACCTCTTTAGCCCGACAAGG  
E G A K S S S V V L S P S T S L A R Q G

1741 CAGTCTGGAGTCACCGTCGTCGGGTACGGGAAGCATGGGCAGTGCTGGTGGGCTGAGTGG  
S L E S P S S G T G S M G S A G G L S G

1801 CAGCAGCAGCCCTCTCTCAATAAACCCCTCAGACCTAACTACAGATGTTATAAGCTTAAG  
S S S P L F N K P S D L T T D V I S L S

1861 TCACTCCTTGGCTTCCAGCCCAGCGTCGGTTCACCTCTTTCACATCCGGTGGGCTTGTGTG  
H S L A S S P A S V H S F T S G G L V W

1921 GGCTGCCAATCTGAGCAGTTTCTCTGCCGGCAGCAAGGACACTCCAAGTTACCAGTCCAT  
A A N L S S S S A G S K D T P S Y Q S M

1981 GACTAGTCTCCATACGAGCTCTGAGTCCATTGACCTGCCCTCAGCCATCATGGCTCCCT  
T S L H T S S E S I D L P L S H H G S L

2041 GTCTGGACTGACCACAGGCACTCACGAGGTGCAGAGCCTGCTCATGAGAACGGGTAGTGT  
S G L T T G T H E V Q S L L M R T G S V

2101 GAGATCTACTCTCTCAGAAAGATACACCCCATCATCTCGGCAGGCCAACCAAGAAGAAGG  
R S T L S E R Y T P S S R Q A N Q E E G

2161 CAAAGAGTGGCTGCGATCGCATTCCTGCGGGGCTGCAGGATACTGGCAACCAGTCTCC  
K E W L R S H S T G G L Q D T G N Q S P

2221 CTGGTCTCCCCCTCTGCCATGTCATCGTCAGCCACCGGAAAATATCACTTTTCCAACCTT  
L V S P S A M S S S A T G K Y H F S N L

2281 GGTGAGTCCCACCAACCTCTCCAGTTTAACCTGCCTGCACCCAGTATGATGCGCTCCAG  
V S P T N L S Q F N L P A P S M M R S S

2341 CAGTATCCCCGCCAGGACTCCTCCTCGACCTCTATGATGATGCCAGCTTTGCGGTAG  
S I P A Q D S S F D L Y D D A Q L C G S

2401 TGCAACTTCCCTGGAGGAAAGGCCACGGCCGTTAGCCACTCCGGCTCATTCAGAGACAG  
A T S L E E R P R A V S H S G S F R D S

2461 CATGGAGGAAGTTCATGGCTCTTCACTGTGCTATTGGTCTCCAGCACATCATCCCTTTACTC  
M E E V H G S S L S L V S S T S S L Y S

2521 TACGGCTGAAGAGAAGGCTCATTCAGAGCAAATCCATAAGCTACGGAGAGAACTGGTTGC  
T A E E K A H S E Q I H K L R R E L V A

2581 CTCCCAGGAGAAAGTCGCTACCCCTACGTCTCAGCTGTCAGCAAATGCTCACCTTGTAGC  
S Q E K V A T L T S Q L S A N A H L V A

2641 AGCTTTTGAAGAGTGTAGGGAATATGACTGGCCGTTTGCAAAGTCTAACCATGACAGC  
A F E K S L G N M T G R L Q S L T M T A

2701 GGAACAAAAGGAATCTGAGCTTATCGAACTGCGGGAAACCATTGAAATGTTGAAGGCCCA  
E...Q K E S E L I E L R E T I E M L K A Q

Fig. 9 (cont'd 1)

68/124

2761 GAACTCTGCTGCCCCAAGCAGCCATTGAGGGAGCACTGAATGGCCCAGACCACCCTCCCAA  
N S A A Q A A I Q G A L N G P D H P P K

2821 AGATCTCCGCATCAGAAGACAGCACTCCTCTGAAAGTGTTCTAGTATCAACAGCGCAAC  
D L R I R R Q H S S E S V S S I N S A T

2881 GAGCCATTCCAGCATTGGCACTGGTAATGATGCTGACTCCAAGAAA  
S H S S I G S G N D A D S K K

Fig. 9 (cont'd 2)

69/124

## Murine genomic sequence

1 GGGATGAAGG GAAAAAGCCA CCGTCAGGCA TTGGAAGATC GACAGCCAGC  
51 AGTTCTTTTG GATACAAGAA GCCAAGTGGT GTAGGGGCTT CCACTATGAT  
101 TACCAGCAGC GGTGCCACCA TCACAAGCGG TTCAGCTACA CTGGGGAAAA  
151 TCCCCAAATC CGCTGCCATT GGTGGGAAGT CCAATGCAGG AAGGAAAACC  
201 AGCCTGGACG GGTCCCAGAA TCAAGATGAT GTTGTCTTGC ACGTGAGCTC  
251 GAAGACCACC CTCCAGTACC GTAGTTTGCC CCGCCCTTCT AAGTCCAGCA  
301 CCAGCGGAAT CCC TGGGAGA GGTGGCCACA GGTGAGCAC CAGCAGCATT  
351 GATTCCAATG TCAGCAGCAA GTCAGCTGGG GCCACCACCT CCAAAGTGA  
401 AGAACCGACT AAGATCGGCT CAGGGCGCTC GAGTCCAGTC ACTGTCAACC  
451 AAACAGACAA AGAGAAGGAG AAAGTAGCAG TGTCAGATTC AGAGAGCGTT  
501 TCCTTGTCAG GTTCCCCCAA ATCCAGCCCC ACCTCTGCCA GTGCCTGTGG  
551 GACTCAAGGG CTCAGACAGC CAGGGTCCAA ATATCCAGAT ATTGCCTCGC  
601 CCACATTTTCG AAGGTAAGGG TATGTAAAGA GATGTTGGGA AAACATAAAA  
651 GG TAGTATAT AGCATGTATT TATTCTGTAC GAAACTATTT TCATGTATTC  
701 TAAATATTCT AAGATTCTGT ATCTTATACT TGTCTAAAAT ATAGTGATTT  
751 TATTTTGCTG ATTGCACCTG TTGCTAGTGT AAAAGCATTG CTCATTTAGA  
801 GAGTGGTTAG CCTTTCAGCT ATACAGCCAG TGTGACACTA AAATACAGAT  
851 ACCACTTGTA GCGGGCATAA AACCACATGA CTGACTATTC ATAGAAATAA  
901 AGTGATAGCT TGTAAGATA TTTAGTGATT TCCACCTCTC CTTTCCAGAA  
951 TTAATAAAG CAAATTGCAT AGATCTTTAT AAACACATTT ACTTCTAGTG  
1001 TATGTTATCT TGTTGACTCT TAATGAAATG GCAGTTATGA ATATAGATGA  
1051 TATATTCTTT CTAACAGTTT ATAAGAGACC AATTTATACA GTACCAGATC  
1101 TTAACATAGT AACAATAACA GCAACAAAAA CAACCCAAAA AGCTATCAAA  
1151 GTATGGTCTG ATTGCAGAAT TTGAAAACAT TTACATGTTT GACATAGGAC  
1201 AAGAACTCAG GAGTGAGGTG ACTTTTATA AGTCTTCATC AATGTCCTTT  
1251 TACAGGAACC AGGAAGCATA TCTGATATAT GTGTCAGGAT TATCACTTTA  
1301 TTAATTATGT GAAATTCTGT TTAGAAATCT ACCTGATTTT AAATACTTTA  
1351 ATATAGTAGG GGTCAAAATT AGTTAATGAG TTAAGACAAG TTGTAAATA  
1401 ATCCTGGCTC TGTTTTCTCA TCTTCAAAAT GATAGAGTAT AATTTATCAC  
1451 CTCTGTGTTA ATATTTCAGG TTTGTGTTTA TTCTCTTGAT AACTTTGATC  
1501 TCTTAGAAGA GTC TTGAAGA ATTTACATTA AGTAATCTTA GAAACATAAC  
1551 TATTTGAGAA ACAGTAGTCA AATTTTGTCA TTAGAAGTAT TAACTCTGAA  
1601 GAATGATTTG AAGTGACAGT TCTTAGAAAG AATAAATTAT AGCTTGTAGC  
1651 AAGAGTAAAT ATTTTCACTG CTTGTGTGAG AGCCAAGAGC GCCCTCTTGT  
1701 GGCCCATAC CTATGAAACA ATTTCTCATA TTCGCCCTAG AAATCTTCCA

Fig. 10

70/124

1751 CTGCAGGAAA TAATGGATTT CATTGCCTCT GAATTAGTAA CCATTCTGCC  
1801 ATTTCTTCAT ACCATTTTAT TTCCATACTT GCATAAATTT GATTATGTCA  
1851 TCTGCTTCAT TTACAAAAC AAAATGTTTT CTGAGCTAAA CTCCAGTAGC  
1901 TAACTTAGTA CAAATGGTAT TTTTAAATCA CTGCTATAAG TATATATATT  
1951 TGAATAGCTC TGGCAACGGA CGGAAATCCC TATGGTCTTT CCATGGGAAG  
2001 ATACAAACCA ATCCATAAGT TGTCCAGCAA TATCCAATAT TTCCAGCCCA  
2051 GCCAGTCAGG CCTCTTAAAC ATTACCTTAC ATATTTGAAC CTTTCCTTAA  
2101 ATGTCCCCTT TAGACAATCT ATTTTTTAAA AAGATGAAAA TCCATTTAAG  
2151 CATCATATAT CGAATGCGTA GAAGTTGTTT CATTATAATG GTTCTGCAGA  
2201 TAGGTAATGC CAAAACGGCC AAAATATTTG ATCACTAGAA GCGTAAAAGT  
2251 CAAGTACAAT CATGTTGACT TTTTTTCCAA GGTGGGTTC A CTGCTGCCCCA  
2301 CCTTGGTTCC AGGCCAGTGC TTA CTTAAGA TATCGTAAGT GATTTTTTTT  
2351 TAATTTTTTAA TTTTTTAGTA GTTGGTTAAT CAAAAGCCAG TCATGTCACC  
2401 TTCAGGAACA TAGAGGCTGG ACGTGCTTGG CAGCTCACGA CTCCAAAGCA  
2451 CACTTGGCTC TGTGGACTGA AACCTAGGA AACGTGGATG TGAGTCTCTT  
2501 GGAACAAC TC AAGTTGTTAT TTGTTTTTCT TTTAGGTTGT TCGGTGCCAA  
2551 GGCAGGCGGC AAATCTGCCT CCGCACCTAA TAC

Fig. 10 (cont'd)



71/124

T2HC

Homologous human cDNA

1 GGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGTGGACCTGCTGAAAGCAGA  
D Q L R E T M H N M Q L E V D L L K A E

61 GAATGACCGACTGAAGGTAGCCCCAGGCCCTCATCAGGCTCCACTCCAGGGCAGGTCCC  
N D R L K V A P G P S S G S T P G Q V P

121 TGGATCATCTGCATTATCTTCCCCACGCCCTCCCTAGGCCTGGCACTCACCCATTCTT  
G S S A L S S P R R S L G L A L T H S F

181 CGGCCCCAGTCTTGCAGACACAGACCTGTCACCCATGGATGGCATCAGTACTTGTGGTCC  
G P S L A D T D L S P M D G I S T C G P

241 AAAGGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCCGCAGCACATCATCAAAGG  
K E E V T L R V V V R M P P Q H I I K G

301 GGACTTGAAGCAGCAGGAATTCTTCTGGGCTGTAGCAAGGTCAGTGGAAGGTTGACTG  
D L K Q Q E F F L G C S K V S G K V D W

361 GAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAAGGACTATATTTCTAAAATGGACCC  
K M L D E A V F Q V F K D Y I S K M D P

421 AGCCTCTACCCTGGGACTAAGCACTGAGTCCATCCATGGCTACAGCATCAGCCACGTGAA  
A S T L G L S T E S I H G Y S I S H V K

481 ACGAGTGTGGATGCAGAGCCCCCGAGATGCCTCCTTGCCGTCGAGGTGTCAATAACAT  
R V L D A E P P E M P P C R R G V N N I

541 ATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTCGACAGCCTGGTGTTCGAGACGCT  
S V S L K G L K E K C V D S L V F E T L

601 GATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCTGAAGCACCGGCGCCTCGT  
I P K P M M Q H Y I S L L L K H R R L V

661 CCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAATCGCTTGGCCGAGTACCT  
L S G P S G T G K T Y L T N R L A E Y L

721 GGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAGCACCTTCAACATGCACCA  
V E R S G R E V T E G I V S T F N M H Q

781 GCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGCCAACAGATAGACCGGGA  
Q S C K D L Q L Y L S N L A N Q I D R E

841 AACAGGAATTGGGGATGTGCCCTGGTGATTCTATTGGATGACCTGAGTGAAGCAGGCTC  
T G I G D V P L V I L L D D L S E A G S

901 CATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCATAAATGTCCCTATATTAT  
I S E L V N G A L T C K Y H K C P Y I I

961 AGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACCTTGAGCTTCAG  
G T T N Q P V K M T P N H G L H L S F R

1021 GATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTTCTGGTTTCGTTACCTGAG  
M L T F S N N V E P A N G F L V R Y L R

1081 GAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAAGGAAGAGCTGCTTCGGGT  
R K L V E S D S D I N A N K E E L L R V

1141 GCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTTCCTTGAGAAGCACAGCAC  
L D W V P K L W Y H L H T F L E K H S T

Fig. 11

72/124

1201 CTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCGTGTCCTTGGCATTGAGGACTT  
S D F L I G P C F F L S C P I G I E D F

1261 CCGGACCTGGTTTCATTGACCTGTGGAACAACCTCTATCATTCCTTATCTACAGGAAGGAGC  
R T W F I D L W N N S I I P Y L Q E G A

1321 CAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGAGGACCCAGTGGAATGGGT  
K D G I K V H G Q K A A W E D P V E W V

1381 CCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATCAAAGCTGTACCACCTGCC  
R D T L P W P S A Q Q D Q S K L Y H L P

1441 CCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGAGGATAGGACAGTCAAAGA  
P P T V G P H S I A S P P E D R T V K D

1501 CAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCATGCTGCTGAAACTTCAAGA  
S T P S S L D S D P L M A M L L K L Q E

1561 AGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCTGGACCCCAACCTTCAGGC  
A A N Y I E S P D R E T I L D P N L Q A

1621 AACACTTTAAGGGTTCGGCAATCAC'TGTACCCCCGGACAGCAGAACGCTGGCATCAGCT  
T L \*

1681 ATCTTAGCTCCTCCTCTCCCCCTCTCCTCTTTCAGAGCACTGGCTCTCCAGCCCCAGGAGG

1741 AGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTC'TTGGTGCTGTACCTTTGA

1801 GAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAAC'TTGTGCCCCCTAAACACATT

1861 TACTGGCCTCCTCTAATGACTTTTGGGGAAAAGATGATTCTGGGTCTTTCCCTTGACTTCT

1921 TGTTTCAATTACAAACTCCTGGGCTTTC'TGGGGAGGGGTTCAGAAAACATCAAAACACTG

1981 CAGCAGTTCCTAAATGATTCTCACAAGCAACCC'TGAGAGAGACAGTCTTGTGAGGGAGAT

2041 CTGGGGGAGGCAGGAAGCTCCTCAGATTTTCTCACAGACCCTTCCCAATTCCATCACCAC

2101 TGCCAACAAC'TCCTCCCCCAGAGATCTGGCTGGAGCCCAGAAAAAGAAGCATGTGGTTTA

2161 AAAAAATGTTTAAATCAATCTGTAAAAGGTAAAAATGAAAAACAAAAACAAGCAAAACAAAC

2221 AAAAAACAATGGAAAAGATGAAGCTGGAGAGAGAGGAACCAGTTGCCAAGGTAGAGAGCT

2281 GCCCGCTCCTGCCCTCTGGATGACATAGGGGACATCAACAAGACGGCTGCCAACCTGAGA

2341 AGTCACCAAACCACAAAAATAACCTTACAGCCTTCAGGGAAAGACTACCAGCTCTGTCTT

2401 TCTACCCCTAATTTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAA

Fig. 11 (cont'd)

73/ 124

Homologous murine cDNA sequence

1 GAACTATGGGAAAAAGAGATGAAGCTCACGGATATCCGGTTGGAGGCCCTCAACTCTGCC  
E L W E K E M K L T D I R L E A L N S A

61 CACCAGCTGGACCAGCTTCGGGAGACCATGCACAATATGCAGTTGGAGGTGGACCTGCTG  
H Q L D Q L R E T M H N M Q L E V D L L

121 AAAGCAGAGAATGACCGGCTGAAGGTTGCCCCCGGCCCTCCTCAGGCTGCACTCCAGGG  
K A E N D R L K V A P G P S S G C T P G

181 CAGGTCCCTGGGTCATCGGCTCTGTCGTCCCCTCGACGTTCCCTGGGCCTTGCACTCAGC  
Q V P G S S A L S S P R R S L G L A L S

241 CATCCTTTTCAGTCTCTAGTCTCACAGACACAGACCTCTCACCCTATGGATGGCATCAGCACC  
H P F S P S L T D T D L S P M D G I S T

301 TGTGGTTCAAAGGAAGAGGTGACCCCTGCGGGTGGTGGTCCGGATGCCGCCCCAGCACATC  
C G S K E E V T L R V V V R M P P Q H I

361 ATCAAAGGGGACTTAAAGCAGCAGGAGTTCTTCTGGGTTGCAGCAAGGTCAGTGGCAAA  
I K G D L K Q Q E F F L G C S K V S G K

421 GTTGACTGGAAGATGCTGGATGAAGCCGTTTTCCTCAAGTGTTCAGGACTACATTTCTAA  
V D W K M L D E A V F Q V F K D Y I S K

481 ATGGACCCAGCCTCAACCCTGGGACTGAGCACTGAGTCCATACATGGCTATAGCCTCAGC  
M D P A S T L G L S T E S I H G Y S L S

541 CACGTGAAACGAGTGCTGGATGCTGAGCCCCCAGAGATGCCTCCTTGCCGCGGAGGTGTC  
H V K R V L D A E P P E M P P C R R G V

601 AATAACATATCAGTCGCTCTCAAAGGTCTGAAAGAGAAGTGTGTCGACAGCCTGGTGTTC  
N N I S V A L K G L K E K C V D S L V F

661 GAGACGCTTATCCCAAGCCCATGATGCAGCACTACATCAGCCTCCTGCTCAAGCACCGG  
E T L I P K P M M Q H Y I S L L L K H R

721 CGCCTGGTGTCTCCGGCCCCAGTGGCACCGGCAAGACCTACTTGACCAATCGGCTAGCC  
R L V L S G P S G T G K T Y L T N R L A

781 GAGTACCTGGTGGAGCGCTCCGGCCGCGAGGTACCGGATGGCATCGTCAGCACTTTCAAC  
E Y L V E R S G R E V T D G I V S T F N

841 ATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTACCTCTCCAACCTAGCCAACCAGATA  
M H Q Q S C K D L Q L Y L S N L A N Q I

901 GACCGGGAACAGGGATAGGGGATGTGCCCTTGGTGATCCTCCTGGATGATCTGAGTGAA  
D R E T G I G D V P L V I L L D D L S E

961 GCAGGCTCCATCAGTGAGCTGGTCAATGGGGCCCTCACCTGCAAGTATCACAAATGTCCC  
A G S I S E L V N G A L T C K Y H K C P

1021 TACATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGGCTTGCACTTG  
Y I I G T T N Q P V K M T P N H G L H L

1081 AGCTTCAGGATGCTGACCTTCTCGAACAATGTGGAACCAGCCAATGGCTTTCTGGTCCGT  
S F R M L T F S N N V E P A N G F L V R

1141 TACCTGCGGAGGAAGTTGGTAGAGTCAGACAGTACGTCATGCTAACAAGGAAGAGCTG  
Y L R R K L V E S D S D V N A N K E E L

1201 CTTCCGGTGTCTGGACTGGGTGCCCAAGCTGTGGTATCACCTCCACACCTTCCTGGAGAAG  
L R V L D W V P K L W Y H L H T F L E K

1261 CACAGCACCTCGGACTTCCTCATTTGGCCCTTGCTTCTCCTGTCTGTCCCATTTGGCATC  
H S T S D F L I G P C F F L S C P I G I

1321 GAGGACTTCCGGACCTGGTTTCATTGACCTGTGGAACAATTCATCATCCCCCTATCTACAG  
E D F R T W F I D L W N N S I I P Y L Q

Fig. 12

74/ 124

1381 GAAGGAGCCAAGGATGGGATCAAGGTTTCATGGACAGAAAGCTGCTTGGAAGACCCGGTG  
E G A K D G I K V H G Q K A A W E D P V

1441 GAATGGGTCCGAGACACTCTTCCCTGGCCGTCGGCCCAACAAGACCAATCAAAGCTCTAC  
E W V R D T L P W P S A Q Q D Q S K L Y

1501 CACCTGCCCCCGCCTTCTGTGGGCCCCCACAGCACTGCCTCACCCCGGAGGACAGGACA  
H L P P P S V G P H S T A S P P E D R T

1561 GTCAAAGACAGCACTCCAACTCCCTCGACTCAGATCCCCTGATGGCCATGCTACTGAAA  
V K D S T P N S L D S D P L M A M L L K

1621 CTCCAAGAAGCTGCCAACTACATTGAGTCACCAGATCGAGAGACTATCCTGGACCCCAAC  
L Q E A A N Y I E S P D R E T I L D P N

1681 CTCCAGGCGACACTCTGAGGGCCCGCAGTCACTGTCACCCTGGAGGGCAGAAGGCTGGC  
L Q A T L \*

1741 TTCAGCATCATTAGCTCTCCTCTGCCCTCTTCCTTCATAGCTCTGGCTCACCAGCCTCGC

1801 CAAGAGAACAGGAGGGAAGAAGAGGGCAGGAGGAGGGATGGGTTCTCGGTGCTGAACCTT

1861 TGAGAACTTCTACTAGGAATTGGAGGGGTGGAGTTTGAGAACTCCGTGCCCCTTAACT

1921 ACATTGCTGGCCTCCTCTTACGACTTAGGAGAAAAGATGATTCTGGTCTTTTCTTCAAG

1981 TTTTGTTTACCTACAAACTCTTGGGCTTTCTGGGGAGGGATTTCGGAAGATATAAACAGA

2041 CAAACAAAAACAAACAAACCAACTACAGCAGTTCCAAGCTCGTTCTCACAAACACCTCTG

2101 AGACAGTCACATGTGGGCAAATCTAAGGGAGGCAGGAAGCTCTACAGACTTTCTTGCAAA

2161 CCCTTCCCAGTTCTGTGCACTGCCAACACCTCCCCGCCAGAGACCTGGCCAGAGCCA

2221 AGAAAAGAGAAGCATGTGGTTTAAACAGAAAAACAAAACAAAACAAAAAATATATG

2281 TGTAAATCAACCTGTAGAAGGTAAAAACGGCAATGGAAAAGATGAAGCTGGAAGGAGGGG

2341 CCCAGTTGCCAAGATGGAACGAGAGCTGCCAGATCTTGCCTTCTGGATGACAAGAGGGGA

2401 CATTGCAAGATGGCTGCCAGTCTAAAACGTCACCAGACCACAAGAGTAACATCACAGCCT

2461 TCGAAGAAAGGCCACAAGCTGTCTTTCTGCCCTCTAACTGAACATGCATGAAAAGTCAAT

2521 AAACCCTACTTTTTTAATTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA

Fig. 12 (cont'd)

## T2 Murine cDNA with following intron

```
CCAATAGAACTCCGGATCAAGAGGCAGAATTCCTCAGATAGCATCTCCAGCCTCAACAGC
1  -----+-----+-----+-----+-----+-----+ 60
a  P I E L R I K R Q N S S D S I S S L N S -
ATCACCAGCCATTCCAGCATCGGCAGCAGCAAAGATGCTGATGCCAAGAAGAAAAAGAAG
61  -----+-----+-----+-----+-----+-----+ 120
a  I T S H S S I G S S K D A D A K K K K K -
AAGAGTTGGGTATGTAAAGGCTTGGGGATCGGCCTGTGCTAGGAGTCACTCACCTGTTG
121  -----+-----+-----+-----+-----+-----+ 180
a  K S W
CAGGGAAGTGAACCCCTTCAGGATCAACAAAGAGGGTCCCTTCTAACAGGATGCCAGTGT
181  -----+-----+-----+-----+-----+-----+ 240
TGTGACATCTGCTGGGGACAAAAATTCATAAGTTCCCATTCTCTATCCATTGTCTATT
241  -----+-----+-----+-----+-----+-----+ 300
CTCCTTACCACCGCCCTGCACATATACCCAGCCCCCACCCTGCCCTGCATCCTTTATAC
301  -----+-----+-----+-----+-----+-----+ 360
ATGTCTGCTATCCTGGGGCTCTACCTACTGATGAGGTCAAATGTATTTGGCCGTAGAAGG
361  -----+-----+-----+-----+-----+-----+ 420
AGCTGAGAAAAATTATTCATGGGTGGGAGAGTGGGGCATGTGGAGAGAATTTGTAAGCCAA
421  -----+-----+-----+-----+-----+-----+ 480
GCAGGGTACTCTAGACGCTCCTGGGGCTGTTGCTTTAGTTTGGGTGAGGAGGCTGTGGAA
481  -----+-----+-----+-----+-----+-----+ 540
CGTCCCCATCGCTCCAAAGCCTGCTTTTGTCTGGTCCAGAGGTGGGTTTGTCTGTGTGG
541  -----+-----+-----+-----+-----+-----+ 600
TATCCCCCTGTAAGTCTAACTGGCTTTGGGTGAGCTTTCTACAATCTGTACGCAGGTG
601  -----+-----+-----+-----+-----+-----+ 660
TAGGGCACTGCCTGACTGACTGAAAGGGAGAGTGACCCAGAGTGAGCGTCTTGTCCTGT
661  -----+-----+-----+-----+-----+-----+ 720
CCCTGCTGAGGAGGGCTGGCTACAGACTTTGGCCTAGTGCAGACAGGAGCCAGCTGTGTG
721  -----+-----+-----+-----+-----+-----+ 780
GAGAAGCAGCTGTGTGAAATGCATGAGTAGTGTGCTGCTGCTGCTGCTGCTGCTTTCTT
781  -----+-----+-----+-----+-----+-----+ 840
TTCATTGTTTTTTTTTTTTTTTCTTTCCCTTTTATTTCCCTTCAAAATGCTGACCTCAAATC
CCTATTTTTTTTCCAGGTTTATGAGGTAAGAACTCGGTTCCCTCTCCTCGTGCTTTTTCT
901  -----+-----+-----+-----+-----+-----+ 960
TTCCCTTTGCACACCTTCGTGTTTCCAGAGCAAGCACCTCTCTTAAAAAAAAAAAAAAAAA
961  -----+-----+-----+-----+-----+-----+ 1020
AAAAA
1021 ----- 1025
```

Fig. 13

76/124

splicing variant 1 (JFC410)

1 AGCGAGTTACTCACGCTTCCCCCTCCATCGGAAGCCAGCCAGGCCAAAACCCAGCAAGATA  
R V T H A S P P S E A S Q A K T Q Q D M

61 TGCAGTCCAGTCTGGCAGCCAGATATGCAACTCAGTCTAATCACAGTGGAATTGCAACCA  
Q S S L A A R Y A T O S N H S G I A T S

121 GTCAAAAAAGCCTACTAGGCTTCCAGGGCCCTCTAGGGTGCCTGCTGCAGGAAGCAGCA  
O K K P T R L P G P S R V P A A G S S S

181 GCAACCTCCACCCACCCCTCTAATTTAAATAGGAGAAGTCAGAGCTTTAACAGCATTGACA  
K V Q G A S N L N R R S Q S F N S I D K

241 AA

bp 1 corresponds to bp 914 of THC

underlined sequence represents further splicing form and is not shown in the THC sequence

Fig. 14

77/124

splicing variant 2

```
1      GGCACCTCAGAGGTCCAGAGCCTGCTCATGAGAACGGGTAGTGTGAGATCTACTCTCTCA
      G T H E V Q S L L M R T G S V R S T L S

61     GAAAGATATACCCCATCATCTCGGCAGGCCAACCAAGAAGAGGGCAAAGAGTGGTTGCGT
      E R Y T P S S R Q A N Q E E G K E W L R

121    TCTCATTCTACTGGAGGGCTTCAGGACACTGGCAACCAG
      S H S T G G L Q D T G N Q
```

bp 1 corresponds to bp 3300 of THC

underlined base pairs -> position of the differentially spliced exon which  
lacks here but is shown in the THC sequence

Fig. 15

## T2-cDNA sequence and T2 protein encoded therein

```

CCGCGGGGCTTCCATCCTTCCTTTGACTGATTTTTAAATTTTAATTTGTATTTTCCCCGC
1  -----+-----+-----+-----+-----+-----+ 60
   R G A S I L P L T D F * I L I C I F P A -

CGCCCCGCCCCCTTTTCCTCCGACCCCGCCCTATCGCTCCCCGGCTTCCCTGCTCTTTCT
61  -----+-----+-----+-----+-----+-----+ 120
   A P P L F L R P R P I A P R L P C S F L -

TTTTCCCGGCTTCCCTTCCTCGCGTTTCTTTCCCTGCGCCCTCGGCTTGCCTCTCTCCCT
121 -----+-----+-----+-----+-----+-----+ 180
   F P G F L P R V S F P C A L G L P L S L -

CCTCCCTCGCTCTCTCCCCCTTCTCTCCCCTTCTTCCTCGGTTTCTTCCGTCCTCTCTCT
181 -----+-----+-----+-----+-----+-----+ 240
   L P R S L P L L S P S S S V S S V L S L -

CCCCCTCCTCCTCCCCCGCCTCCTCCTCCTGCGCTCCCGCCCCCTGCCCTCCCCCGT
241 -----+-----+-----+-----+-----+-----+ 300
   P L L L P R L L L L R S R P L P P P P V -

GCCTGCAGACGCGCGGATCGTCCATGCGCTCCTCGCGGGCAGAATGCTGGGCAGCAGCGT
301 -----+-----+-----+-----+-----+-----+ 360
   P A D A R I V H A L L A G R M L G S S V -

CAAGAGCGTGCAGCCCCGAGGTGGAGCTGAGCAGCGGCGGCGGACGAGGGCGCGGACGA
361 -----+-----+-----+-----+-----+-----+ 420
   K S V Q P E V E L S S G G G D E G A D E -

ACCGCGGGGCGCCGGCAGGAAGGCGGCAGCGGCGGACGGCAGAGGCATGCTGCCCAAGCG
421 -----+-----+-----+-----+-----+-----+ 480
   P R G A G R K A A A A D G R G M L P K R -

CGCCAAGGCGCCCGGCGGCGGCGGCGGCATGGCCAAGGCCAGCGCGGCTGAGCTGAAGGT
481 -----+-----+-----+-----+-----+-----+ 540
   A K A P G G G G G M A K A S A A E L K V -

CTTCAAGTCCGGCAGCGTGGACAGCCGTGTCCCCGGCGGGCCCGCCCTCCAACCTGCG
541 -----+-----+-----+-----+-----+-----+ 600
   F K S G S V D S R V P G G P P A S N L R -

CAAGCAGAAGTCACTCACCAACCTCTCTTTTCTCACGGACTCCGAGAAAAAGCTGCAGCT
601 -----+-----+-----+-----+-----+-----+ 660
   K Q K S L T N L S F L T D S E K K L Q L -

TTATGAGCCCGAATGGAGCGACGATATGGCCAAGGCGCCCAAAGGCTTAGGCAAGGTGGG
661 -----+-----+-----+-----+-----+-----+ 720
   Y E P E W S D D M A K A P K G L G K V G -

GTCCAAGGGCCGTGAAGCTCCGCTGATGTCCAAGACGCTGTCCAAGTCGGAGCACTCGCT
721 -----+-----+-----+-----+-----+-----+ 780
   S K G R E A P L M S K T L S K S E H S L -

CTTCCAGGCCAAGGGCAGCCCGGCGGGCGGCCAAGACCCCCCTGGCTCCGCTCGCGCC
781 -----+-----+-----+-----+-----+-----+ 840
   F Q A K G S P A G G A K T P L A P L A P -

```

Fig. 16



CAACCTGGGAAAGCCGAGCCGGATCCCTCGAGGACCCTATGCGGAGGTCAAGCCGCTCAG  
841 -----+-----+-----+-----+-----+ 900  
N L G K P S R I P R G P Y A E V K P L S -

CAAGGCGCCTGAAGCGGCCGTGAGCGAAGATGGCAAATCGGACGACGAGCTGCTCTCCAG  
901 -----+-----+-----+-----+-----+ 960  
K A P E A A V S E D G K S D D E L L S S -

CAAGGCCAAGGCGCAAAAGAGCTCTGGGCCTGTCCCCTCTGCCAAGGGCCAGGAGGAGCG  
961 -----+-----+-----+-----+-----+ 1020  
K A K A Q K S S G P V P S A K G Q E E R -

CGCCTTCCTCAAGGTGGACCCCGAGCTGGTGGTGACCGTGCTGGGAGACCTGGAGCAGCT  
1021 -----+-----+-----+-----+-----+ 1080  
A F L K V D P E L V V T V L G D L E Q L -

GCTCTTCAGCCAGATGCTGGACCCAGAGTCCCAGAGAAAGAGGACAGTGCAGAATGTCCT  
1081 -----+-----+-----+-----+-----+ 1140  
L F S Q M L D P E S Q R K R T V Q N V L -

GGATCTCCGGCAGAACCTGGAAGAGACCATGTCCAGCCTGCGAGGGTCCCAGGTGACTCA  
1141 -----+-----+-----+-----+-----+ 1200  
D L R Q N L E E T M S S L R G S Q V T H -

CAGCTCCCTGGAGATGACCTGCTACGACAGCGATGATGCCAACCCACGCAGCGTGCTCCAG  
1201 -----+-----+-----+-----+-----+ 1260  
S S L E M T C Y D S D D A N P R S V S S -

CCTCTCCAACCGCTCGTACCCTCTGTCTATGGCGCTATGGCCAGTCCAGTCCGCGGCTGCA  
1261 -----+-----+-----+-----+-----+ 1320  
L S N R S Y P L S W R Y G Q S S P R L Q -

GGCTGGTGACGCGCCCTCTGTGGGTGGGAGCTGCCGCTCGGAGGGGACGCCCGCCTGGTA  
1321 -----+-----+-----+-----+-----+ 1380  
A G D A P S V G G S C R S E G T P A W Y -

CATGCACGGCGAACGGGCCCCTACTCCCACACCATGCCCATGCGCAGCCCCAGCAAGCT  
1381 -----+-----+-----+-----+-----+ 1440  
M H G E R A H Y S H T M P M R S P S K L -

CAGCCATATCTCCCGCCTGGAGCTGGTGAATCCCTGGACTCGGATGAGGTGGACCTCAA  
1441 -----+-----+-----+-----+-----+ 1500  
S H I S R L E L V E S L D S D E V D L K -

GTCCGGCTACATGAGCGACAGTGACCTCATGGGCAAGACCATGACGGAGGATGATGACAT  
1501 -----+-----+-----+-----+-----+ 1560  
S G Y M S D S D L M G K T M T E D D D I -

CACTACCGGCTGGGATGAAAGCAGCTCCATCAGTAGTGGACTCAGCGATGCCTCAGACAA  
1561 -----+-----+-----+-----+-----+ 1620  
T T G W D E S S S I S S G L S D A S D N -

TCTCAGTTCAGAAGAATTCAATGCCAGCTCCTCACTCAACTCCCTCCCAAGTACTCCAC  
1621 -----+-----+-----+-----+-----+ 1680  
L S S E E F N A S S S L N S L P S T P T -

TGCTTCTCGCAGGAAGTCAACAATAGTGCTACGCACAGACTCAGAGAAGCGCTCACTGGC  
1681 -----+-----+-----+-----+-----+ 1740  
A S R R N S T I V L R T D S E K R S L A -

Fig. 16 (cont'd 1)

80/124

```

AGAAAGTGGGCTGAGCTGGTTTtagtgaatcagaggagaaagccccctaaaaaactggagta
1741 -----+-----+-----+-----+-----+-----+-----+ 1800
      E S G L S W F S E S E E K A P K K L E Y -

CGACAGTGGTAGCCTGAAGATGGAACCTGGGACTTCTAAGTGGCGGAGGGAGCGGCCTGA
1801 -----+-----+-----+-----+-----+-----+-----+ 1860
      D S G S L K M E P G T S K W R R E R P E -

GAGCTGTGATGATTTCATCCAAGGGTGGAGAACTGAAAAAGCCCATCAGCCTGGGCCACCC
1861 -----+-----+-----+-----+-----+-----+-----+ 1920
      S C D D S S K G G E L K K P I S L G H P -

TGGTTCCTGAAGAAGGGCAAGACCCACCTGTGGCTGTAACCTCCCCCATCACTCACAC
1921 -----+-----+-----+-----+-----+-----+-----+ 1980
      G S L K K G K T P P V A V T S P I T H T -

AGCCCAGAGTGCCCTCAAAGTCGCAGGCCAAACCTGAGGGCAAAGCTACAGACAAGGGTAA
1981 -----+-----+-----+-----+-----+-----+-----+ 2040
      A Q S A L K V A G K P E G K A T D K G K -

GCTTGCAGTGAAGAATACTGGGCTCCAACGCTCCTCCTCTGATGCTGGTCGGGACCGCCT
2041 -----+-----+-----+-----+-----+-----+-----+ 2100
      L A V K N T G L Q R S S S D A G R D R L -

GAGTGATGCTAAGAAGCCCCCTCGGGCATTGCTCGCCCCCTCCACTTCGGGATCCTTTGG
2101 -----+-----+-----+-----+-----+-----+-----+ 2160
      S D A K K P P S G I A R P S T S G S F G -

CTACAAGAAGCCTCCTCCTGCCACAGGCACAGCCACTGTGTCAGAACTGGTGGTTCAGC
2161 -----+-----+-----+-----+-----+-----+-----+ 2220
      Y K K P P P A T G T A T V M Q T G G S A -

CACTCTCAGCAAGATCCAGAAGTCCTCAGGCATCCCTGTCAAGCCAGTAAATGGGCGCAA
2221 -----+-----+-----+-----+-----+-----+-----+ 2280
      T L S K I Q K S S G I P V K P V N G R K -

GACTAGCTTAGATGTTTCCAACAGTGCAGAGCCAGGATTCCTGGCTCCTGGAGCCCGTTC
2281 -----+-----+-----+-----+-----+-----+-----+ 2340
      T S L D V S N S A E P G F L A P G A R S -

TAACATCCAGTACCGCAGCCTGCCCCGGCCAGCCAAGTCAAGTTCTATGAGCGTGACCGG
2341 -----+-----+-----+-----+-----+-----+-----+ 2400
      N I Q Y R S L P R P A K S S S M S V T G -

CGGGCGGGGTGGACCTCGCCCTGTGAGCAGCAGCATTGACCCCAGTCTCCTCAGCACCAA
2401 -----+-----+-----+-----+-----+-----+-----+ 2460
      G R G G P R P V S S S I D P S L L S T K -

GCAGGGAGGCCTTACGCCTTCCAGACTGAAGGAGCCTACCAAGGTAGCCAGTGGGCGGAC
2461 -----+-----+-----+-----+-----+-----+-----+ 2520
      Q G G L T P S R L K E P T K V A S G R T -

CACTCCAGCCCCTGTCAATCAGACAGATCGGGAAAAGGAGAAGGCCAAAGCCAAGGCAGT
2521 -----+-----+-----+-----+-----+-----+-----+ 2580
      T P A P V N Q T D R E K E K A K A K A V -

GGCCTTGGACTCAGACAACATCTCCTTGAAGAGTATTGGCTCCCCAGAAAGTACTCCCAA
2581 -----+-----+-----+-----+-----+-----+-----+ 2640
      A L D S D N I S L K S I G S P E S T P K -

```

Fig. 16 (cont'd 2)

81/124

GAACCAAGCAAGCCACCCACAGCCACCAAGCTGGCAGAGCTGCCACCAACCCCTCTCAG  
 2641 -----+-----+-----+-----+-----+-----+-----+ 2700  
 N Q A S H P T A T K L A E L P P T P L R -  
 GGCCACAGCGAAGAGCTTTGTCAAACCACCCCTCACTAGCCAATCTTGACAAGGTCAACTC  
 2701 -----+-----+-----+-----+-----+-----+-----+ 2760  
 A T A K S F V K P P S L A N L D K V N S -  
 CAACAGTCTGGATCTACCATCATCCAGTGATACCACCCATGCTTCAAAGGTCCCAGATCT  
 2761 -----+-----+-----+-----+-----+-----+-----+ 2820  
 N S L D L P S S S D T T H A S K V P D L -  
 GCATGCTACAAGCTCAGCATCTGGGGGCCCTCTCCCTTCCTGCTTCACCCCCAGTCCGGC  
 2821 -----+-----+-----+-----+-----+-----+-----+ 2880  
 H A T S S A S G G P L P S C F T P S P A -  
 ACCCATCCTCAATATTAACCTCAGCCAGCTTCTCCCAGGGCCTGGAGCTAATGAGTGGTTT  
 2881 -----+-----+-----+-----+-----+-----+-----+ 2940  
 P I L N I N S A S F S Q G L E L M S G F -  
 CAGTGTGCCAAAAGAGACCCGCATGTACCCCAAACCTCTCAGGCCTGCACAGGAGCATGGA  
 2941 -----+-----+-----+-----+-----+-----+-----+ 3000  
 S V P K E T R M Y P K L S G L H R S M E -  
 GTCCCTCCAGATGCCAATGAGCCTCCCCAGTGCCTTCCCCAGCAGTACTCCCGTCCCCAC  
 3001 -----+-----+-----+-----+-----+-----+-----+ 3060  
 S L Q M P M S L P S A F P S S T P V P T -  
 CCCACCTGCTCCCCCTGCTGCTCCACAGAAGAAGAGACGGAAGAGCTGACTTGGAGTGG  
 3061 -----+-----+-----+-----+-----+-----+-----+ 3120  
 P P A P P A A P T E E E T E E L T W S G -  
 AAGCCCCAGAGCTGGGCAACTGGACAGTAATCAGCGGGATCGGAACACTCTTCCCAAGAA  
 3121 -----+-----+-----+-----+-----+-----+-----+ 3180  
 S P R A G Q L D S N Q R D R N T L P K K -  
 AGGGCTCAGGTACCAGCTTCAGTCCCAGGAGGAGACCAAGGAGAGGCGACATTCCCATAC  
 3181 -----+-----+-----+-----+-----+-----+-----+ 3240  
 G L R Y Q L Q S Q E E T K E R R H S H T -  
 CATTGGTGGGCTGCCTGAATCCGATGACCAGTCAGAGCTGCCTTCTCCCCCTGCACTTCC  
 3241 -----+-----+-----+-----+-----+-----+-----+ 3300  
 I G G L P E S D D Q S E L P S P P A L P -  
 CATGTCTCTGAGTGCAAAGGGCCAACTTACCAACATAGTGAGTCCCACTGCGGCCACCAC  
 3301 -----+-----+-----+-----+-----+-----+-----+ 3360  
 M S L S A K G Q L T N I V S P T A A T T -  
 \*GCCAAGAATCACCCGCTCCAACAGCATCCCCACCCACGAGGCGGCCTTCGAGCTGTACAG  
 3361 -----+-----+-----+-----+-----+-----+-----+ 3420  
 P R I T R S N S I P T H E A A F E L Y S -  
 CGGCTCCCAAATGGGGAGCACCCCTGTCCCTGGCCGAGAGACCCAAGGGAATGATTCGGTC  
 3421 -----+-----+-----+-----+-----+-----+-----+ 3480  
 G S Q M G S T L S L A E R P K G M I R S -  
 AGGATCCTTCCGAGACCCACGGACGATGTTACGGCTCAGTGCTGTCCCTGGCCTCCAG  
 3481 -----+-----+-----+-----+-----+-----+-----+ 3540  
 G S F R D P T D D V H G S V L S L A S S -

Fig. 16 (cont'd 3)

82/124

```

TGCCTCCTCCACCTACTCCTCAGCTGAGGAGAGGATGCAATCTGAGCAAATCCGGAAGCT
3541 -----+-----+-----+-----+-----+-----+ 3600
  A S S T Y S S A E E R M Q S E Q I R K L -

TCGTAGGGAAGTGAATCATCCCAGGAAAAAGTGGCCACCTTGACGTCTCAGCTTTCTGC
3601 -----+-----+-----+-----+-----+-----+ 3660
  R R E L E S S Q E K V A T L T S Q L S A -

CAATGCTAATCTGGTGGCTGCTTTTGAGCAGAGCCTGGTGAATATGACATCCCGCCTGCG
3661 -----+-----+-----+-----+-----+-----+ 3720
  N A N L V A A F E Q S L V N M T S R L R -

ACACCTGGCAGAGACGGCCGAGGAGAAGGACACTGAGCTGCTGGATTTGCGAGAAACCAT
3721 -----+-----+-----+-----+-----+-----+ 3780
  H L A E T A E E K D T E L L D L R E T I -

AGACTTTCTGAAGAAAAAGAACTCTGAGGCCCAGGCAGTCATTCAGGGAGCCCTTAATGC
3781 -----+-----+-----+-----+-----+-----+ 3840
  D F L K K K N S E A Q A V I Q G A L N A -

CTCAGAAACCACACCCAAAGAACTTCGGATCAAGAGACAAAACCTCCTCAGATAGCATCTC
3841 -----+-----+-----+-----+-----+-----+ 3900
  S E T T P K E L R I K R Q N S S D S I S -

AAGCCTCAACAGCATCACTAGCCATTCCAGCATCGGCAGCAGCAAGGATGCTGATGCGAA
3901 -----+-----+-----+-----+-----+-----+ 3960
  S L N S I T S H S S I G S S K D A D A K -

AAAGAAGAAAAAAGAGTTGGCTTCGAAGTTCCTTCAACAAAGCGTTTCAGTATAAAAAA
3961 -----+-----+-----+-----+-----+-----+ 4020
  K K K K K S W L R S S F N K A F S I K K -

GGGGCCCAAGTCAGCTTCCTCATACTCGGATATAGAGGAGATTGCTACACCCGACTCTTC
4021 -----+-----+-----+-----+-----+-----+ 4080
  G P K S A S S Y S D I E E I A T P D S S -

AGCCCCCTCATCCCCCAAACCTACAGCATGGTTCTACAGAGACTGCTTCACCCCTCCATCAA
4081 -----+-----+-----+-----+-----+-----+ 4140
  A P S S P K L Q H G S T E T A S P S I K -

GTCCTCCACCTCGTCCTCCGTGGGCACTGATGTACCGAGGGCCCTGCTCACCCAGCCCC
4141 -----+-----+-----+-----+-----+-----+ 4200
  S S T S S S V G T D V T E G P A H P A P -

CCCACTAGGCTGTTCCATGCAAATGAGGAGGAGGAGCCAGAGAAGAAGGAGGTATCGGA
4201 -----+-----+-----+-----+-----+-----+ 4260
  H T R L F H A N E E E E P E K K E V S E -

GCTGCGCTCTGAGCTATGGGAGAAGGAAATGAAGCTTACAGACATCCGCTTGAGGGCCCT
4261 -----+-----+-----+-----+-----+-----+ 4320
  L R S E L W E K E M K L T D I R L E A L -

CAACTCTGCCCACCAACTGGATCAGCTTCGGGAGACCATGCACAACATGCAGTTGGAGGT
4321 -----+-----+-----+-----+-----+-----+ 4380
  N S A H Q L D Q L R E T M H N M Q L E V -

GGACCTGCTGGAAGCAGAGAATGACCGACTGAAGGTAGCCCCAGGCCCTTCATCAGGCTC
4381 -----+-----+-----+-----+-----+-----+ 4440
  D L L E A E N D R L K V A P G P S S Q S -

```

Fig. 16 (cont'd 4)

83/124

CACTCCAGGGCAGGTCCCTGGATCATCTGCATTATCTTCCCCACGCCGCTCCCTAGGCCT  
 4441 -----+-----+-----+-----+-----+ 4500  
 T P G Q V P G S S A L S S P R R S L G L -  
 GGCACCTACCCATTCCCTTCGGCCCCAGTCTTGACAGACACAGACCTGTCACCCATGGATGG  
 4501 -----+-----+-----+-----+-----+ 4560  
 A L T H S F G P S L A D T D L S P M D G -  
 CATCAGTACTTGTGGTCCAAAGGAGGAAGTGACCCTCCGGGTGGTGGTGAGGATGCCCCC  
 4561 -----+-----+-----+-----+-----+ 4620  
 I S T C G P K E E V T L R V V V R M P P -  
 GCAGCACATCATCAAAGGGGACTTGAAGCAGCAGGAATTCTTCTGGGCTGTAGCAAGGT  
 4621 -----+-----+-----+-----+-----+ 4680  
 Q H I I K G D L K Q Q E F F L G C S K V -  
 CAGTGGAAGTTGACTGGAAGATGCTGGATGAAGCTGTTTTCCAAGTGTTCAGGACTA  
 4681 -----+-----+-----+-----+-----+ 4740  
 S G K V D W K M L D E A V F Q V F K D Y -  
 TATTTCTAAATGGACCCAGCCTCTACCCTGGGACTAAGCACTGAGTCCATCCATGGCTA  
 4741 -----+-----+-----+-----+-----+ 4800  
 I S K M D P A S T L G L S T E S I H G Y -  
 CAGCATCAGCCACGTGAAACGAGTGTTGGATGCAGAGCCCCCGAGATGCCTCCTTGCCG  
 4801 -----+-----+-----+-----+-----+ 4860  
 S I S H V K R V L D A E P P E M P P C R -  
 TCGAGGTGTCAATAACATATCAGTCTCCCTCAAAGGTCTGAAGGAGAAATGCGTGCACAG  
 4861 -----+-----+-----+-----+-----+ 4920  
 R G V N N I S V S L K G L K E K C V D S -  
 CCTGGTGTTCGAGACGCTGATCCCCAAGCCGATGATGCAGCACTACATAAGCCTCCTGCT  
 4921 -----+-----+-----+-----+-----+ 4980  
 L V F E T L I P K P M M Q H Y I S L L L -  
 GAAGCACCGGCGCCTCGTCTCTCGGGCCCCAGCGGCACGGGCAAGACCTACCTGACCAA  
 4981 -----+-----+-----+-----+-----+ 5040  
 K H R R L V L S G P S G T G K T Y L T N -  
 TCGCTTGGCCGAGTACCTGGTGGAGCGCTCTGGCCGTGAGGTCACAGAGGGCATCGTCAG  
 5041 -----+-----+-----+-----+-----+ 5100  
 R L A E Y L V E R S G R E V T E G I V S -  
 CACCTTCAACATGCACCAGCAGTCTTGCAAGGATCTGCAACTGTATCTTTCCAACCTAGC  
 5101 -----+-----+-----+-----+-----+ 5160  
 T F N M H Q Q S C K D L Q L Y L S N L A -  
 CAACCAGATAGACCGGGAAACAGGAATTGGGGATGTGCCCTGGTGATTCTATTGGATGA  
 5161 -----+-----+-----+-----+-----+ 5220  
 N Q I D R E T G I G D V P L V I L L D D -  
 CCTGAGTGAAGCAGGCTCCATCAGTGAGTTGGTCAATGGGGCCCTCACCTGCAAGTATCA  
 5221 -----+-----+-----+-----+-----+ 5280  
 L S E A G S I S E L V N G A L T C K Y H -  
 TAAATGTCCCTATATTATAGGTACCACCAATCAGCCTGTAAAAATGACACCCAACCATGG  
 5281 -----+-----+-----+-----+-----+ 5340  
 K C P Y I I G T T N Q P V K M T P N H G -

Fig. 16 (cont'd 5)

84/124

```

CTTTCAC TTGAGCTTCAGGATGTTGACCTTCTCCAACAACGTGGAGCCAGCCAATGGCTT
5341 -----+-----+-----+-----+-----+-----+-----+ 5400
      F H L S F R M L T F S N N V E P A N G F -

CCTGGTTCGTTACCTGAGGAGGAAGCTGGTAGAGTCAGACAGCGACATCAATGCCAACAA
5401 -----+-----+-----+-----+-----+-----+-----+ 5460
      L V R Y L R R K L V E S D S D I N A N K -

GGAAGAGCTGCTTCGGGTGCTCGACTGGGTACCCAAGCTGTGGTATCATCTCCACACCTT
5461 -----+-----+-----+-----+-----+-----+-----+ 5520
      E E L L R V L D W V P K L W Y H L H T F -

CCTTGAGAAGCACAGCACCTCAGACTTCCTCATCGGCCCTTGCTTCTTTCTGTCTGTGTC
5521 -----+-----+-----+-----+-----+-----+-----+ 5580
      L E K H S T S D F L I G P C F F L S C P -

CATTGGCATTGAGGACTTCCGGACCTGGTTCATTGACCTGTGGAACAACCTCTATCATTCC
5581 -----+-----+-----+-----+-----+-----+-----+ 5640
      I G I E D F R T W F I D L W N N S I I P -

CTATCTACAGGAAGGAGCCAAGGATGGGATAAAGGTCCATGGACAGAAAGCTGCTTGGGA
5641 -----+-----+-----+-----+-----+-----+-----+ 5700
      Y L Q E G A K D G I K V H G Q K A A W E -

GGACCCAGTGGAATGGGTCCGGGACACACTTCCCTGGCCATCAGCCCAACAAGACCAATC
5701 -----+-----+-----+-----+-----+-----+-----+ 5760
      D P V E W V R D T L P W P S A Q Q D Q S -

AAAGCTGTACCACCTGCCCCACCCACCGTGGGCCCTCACAGCATTGCCTCACCTCCCGA
5761 -----+-----+-----+-----+-----+-----+-----+ 5820
      K L Y H L P P P T V G P H S I A S P P E -

GGATAGGACAGTCAAAGACAGCACCCCAAGTTCTCTGGACTCAGATCCTCTGATGGCCAT
5821 -----+-----+-----+-----+-----+-----+-----+ 5880
      D R T V K D S T P S S L D S D P L M A M -

GCTGCTGAAACTTCAAGAAGCTGCCAACTACATTGAGTCTCCAGATCGAGAAACCATCCT
5881 -----+-----+-----+-----+-----+-----+-----+ 5940
      L L K L Q E A A N Y I E S P D R E T I L -

GGACCCCAACCTTCAGGCAACACTTTAAGGGTTCGGCAATCACTGTCACCCCGGACAGC
5941 -----+-----+-----+-----+-----+-----+-----+ 6000
      D P N L Q A T L * -

AGAACGCTGGCATCAGCTATCTTAGCTCCTCCTCTCCCCTCTCCTCTTTTCAGAGCACTGG
6001 -----+-----+-----+-----+-----+-----+-----+ 6060

CTCTCCAGCCCCAGGAGGAGAACAGGAGGGAGGAGGAGATGAAAGAGGAGGGACAGGTTC
6061 -----+-----+-----+-----+-----+-----+-----+ 6120

TTGGTGCTGTACCTTTGAGAACTTCCTAGGAAGGAATGGTGGGGTGGCGTTTGGGAACTT
6121 -----+-----+-----+-----+-----+-----+-----+ 6180

GTGCCCCCTAAACACATTTACTGGCCTCCTCTAATGACTTTGGGGAAAAGATGATTCTGG
6181 -----+-----+-----+-----+-----+-----+-----+ 6240

GTCTTTCCCTTGACTTCTTGTTTCAATTACAAACTCCTGGGCTTTCTGGGGAGGGGTTC
6241 -----+-----+-----+-----+-----+-----+-----+ 6300

```

Fig. 16 (cont'd 6)

85/124

GAAAACATCAAAACACTGCAGCAGTTCCTAAATGATTCTCACAAGCAACCCTGAGAGAGA  
6301 -----+-----+-----+-----+-----+-----+ 6360  
CAGTCTTGTGAGGGAGATCTGGGGGAGGCAGGAAGCTCCTCAGATTTTCTCACAGACCCT  
6361 -----+-----+-----+-----+-----+-----+ 6420  
TCCAATTCCATCACCCTGCCAACAACCTCCTCCCCAGAGATCTGGCTGGAGCCCAGAA  
6421 -----+-----+-----+-----+-----+-----+ 6480  
AAAGAAGCATGTGGTTTAAAAAATGTTTAAATCAATCTGTAAAAGGTAAAAATGAAAAAC  
6481 -----+-----+-----+-----+-----+-----+ 6540  
AAAAACAAGCAAACAAACAAAAACAATGAAAAGATGAAGCTGGAGAGAGAGGAACCAG  
6541 -----+-----+-----+-----+-----+-----+ 6600  
TTGCCAAGGTAGAGAGCTGCCCGCTCCTGCCCTCTGGATGACATAGGGGACATCAACAAG  
6601 -----+-----+-----+-----+-----+-----+ 6660  
ACGGCTGCCAACCTGAGAAGTCACCAAACCACAAAAATAACCTTACAGCCTTCAGGGAAA  
6661 -----+-----+-----+-----+-----+-----+ 6720  
GACTACCAGCTCTGTCTTTCTACCCTCTAATTTAACAATGCATAAGAGTCAATAAACCCCT  
6721 -----+-----+-----+-----+-----+-----+ 6780  
ACTTTTTTAAAAAAAAAAAAAAAAAAG  
6781 -----+-----+----- 6805

Fig. 16 (cont'd 7)

86/124

T3-cDNA sequence and T3 protein  
encoded therein (protein isoform 1)

```
CAACCAGCCAGAACGCCTGAACTCGCAGGTGCTGCAGGGGCTGCAGGAGCCAGCGGGGGA
1  -----+-----+-----+-----+-----+ 60
   N Q P E R L N S Q V L Q G L Q E P A G E -

GGGGCTCCCGCTGCGGAAGAGCGGCTCGGTGGAACCGGGTTCGATACCCAGATCTACAC
61  -----+-----+-----+-----+-----+ 120
   G L P L R K S G S V E N G F D T Q I Y T -

AGACTGGGCCAATCATTACCTAGCCAAATCCGGCCACAAGCGTCTCATCAGGGATCTCCA
121 -----+-----+-----+-----+-----+ 180
   D W A N H Y L A K S G H K R L I R D L Q -

GCAAGATGTGACAGATGGCGTCCTCCTGGCCCAGATTATCCAGGTTGTGGCAAATGAAAA
181 -----+-----+-----+-----+-----+ 240
   Q D V T D G V L L A Q I I Q V V A N E K -

GATTGAAGACATCAATGGCTGTCCGAAGAACAGATCCCAAATGATTGAAAACATAGATGC
241 -----+-----+-----+-----+-----+ 300
   I E D I N G C P K N R S Q M I E N I D A -

CTGCTTGAATTTCTGCGCAGCTAAGGGAATAAACATCCAGGGGCTGTCTGCAGAAGAGAT
301 -----+-----+-----+-----+-----+ 360
   C L N F L A A K G I N I Q G L S A E E I -

CAGGAATGGAAACCTCAAGGCCATTCTAGGCCTCTTCTTCAGCCTCTCCCGATACAAGCA
361 -----+-----+-----+-----+-----+ 420
   R N G N L K A I L G L F F S L S R Y K Q -

GCAGCAGCAGCAGCCCCAGAAGCAGCACCTCTCCTCACCTCTGCCGCCCCGCGTATCCCA
421 -----+-----+-----+-----+-----+ 480
   Q Q Q Q P Q K Q H L S S P L P P A V S Q -

GGTGGCCGGGGCCCCCTCCAGTGCCAGGCTGGCACCCCTCAGCAGCAGGTGCCAGTCAC
481 -----+-----+-----+-----+-----+ 540
   V A G A P S Q C Q A G T P Q Q Q V P V T -

TCCCCAAGCCCCGTGCCAGCCTCACCAGCCAGCGCCACATCAGCAGTCAAAAGCACAAGC
541 -----+-----+-----+-----+-----+ 600
   P Q A P C Q P H Q P A P H Q Q S K A Q A -

TGAAATGCAGTCCAGACTTCCAGGTCCTACCGCGAGGGTATCCGCTGCAGGCAGCGAGGC
601 -----+-----+-----+-----+-----+ 660
   E M Q S R L P G P T A R V S A A G S E A -

CAAAACACGCGGAGGGTCAACTACTGCTAACAACCGACGCAGCCAGAGCTTTAACAATA
661 -----+-----+-----+-----+-----+ 720
   K T R G G S T T A N N R R S Q S F N N Y -

TGATAAATCCAAACAGTCACCTCCCCACCCCCACCGCCAAGCAGCCACGAGAAAGAGCC
721 -----+-----+-----+-----+-----+ 780
   D K S K P V T S P P P P P S S H E K E P -

TTTGGCAAGTTCAGCCTCCTCCCACCCCCGAATGAGTGACAATGCACCTGCTTCCTTGGA
781 -----+-----+-----+-----+-----+ 840
   L A S S A S S H P G M S D N A P A S L E -
```

Fig. 17



GAGCGGCAGCAGCTCCACCCCTACTAATTGCAGTACCTCCTCGGCCATCCCGCAGCCCGG  
 841 -----+-----+-----+-----+-----+-----+-----+ 900  
 S G S S S T P T N C S T S S A I P Q P G -  
 TGCAGCCACCAAGCCTTGGCGCAGCAAATCCCTCAGCGTGAAGCACAGTGCCACGGTATC  
 901 -----+-----+-----+-----+-----+-----+-----+ 960  
 A A T K P W R S K S L S V K H S A T V S -  
 CATGCTCTCGGTCAAGCCTCCTGGGCCTGAGGCCCCCAGGCCCCACACCTGAAGCCATGAA  
 961 -----+-----+-----+-----+-----+-----+-----+ 1020  
 M L S V K P P G P E A P R P T P E A M K -  
 GCCGGCCCCCAACAATCAGAAGTCCATGCTGGAAAAGCTGAACTTTTCAACAGTAAAGG  
 1021 -----+-----+-----+-----+-----+-----+-----+ 1080  
 P A P N N Q K S M L E K L K L F N S K G -  
 GGGCTCAAAGGCAGGTGAGGGGCGGGGTCCCGGGACACAAGCTGTGAGCGGCTGGAGAC  
 1081 -----+-----+-----+-----+-----+-----+-----+ 1140  
 G S K A G E G P G S R D T S C E R L E T -  
 TCTGCCCAGCTTCGAAGAGAGCGAGGAGCTGGAGGCCGCCAGTCGCATGCTCACCACCGT  
 1141 -----+-----+-----+-----+-----+-----+-----+ 1200  
 L P S F E E S E E L E A A S R M L T T V -  
 GGGCCCTGCTTCCAGCAGCCCCAAGATTGCACTCAAGGGCATTGCCAGAGGACTTTTAG  
 1201 -----+-----+-----+-----+-----+-----+-----+ 1260  
 G P A S S S P K I A L K G I A Q R T F S -  
 CCGGGCACTGACCAACAAGAAGAGTTCTCTGAAAGGCAATGAGAAAGAGAAGGAGAAACA  
 1261 -----+-----+-----+-----+-----+-----+-----+ 1320  
 R A L T N K K S S L K G N E K E K E K Q -  
 ACAGCGGGAGAAGGATAAGGAGAAAAGCAAGGACCTTGCCAAGAGAGCCTCTGTGACGGA  
 1321 -----+-----+-----+-----+-----+-----+-----+ 1380  
 Q R E K D K E K S K D L A K R A S V T E -  
 GAGGCTGGACCTCAAGGAGGAGCCAAAAGAAGACCCAGTGAGCAGCTGTGCCCGAGAT  
 1381 -----+-----+-----+-----+-----+-----+-----+ 1440  
 R L D L K E E P K E D P S G A A V P E M -  
 GCCAAAAAAGTCCTCCAAGATTGCCAGCTTCATCCCCAAAGGGGGGAAGCTCAACAGTGC  
 1441 -----+-----+-----+-----+-----+-----+-----+ 1500  
 P K K S S K I A S F I P K G G K L N S A -  
 CAAGAAGGAGCCCATGGCCCCTTCCCACAGTGAATACCAAACCAGGAATGAAGAGCAT  
 1501 -----+-----+-----+-----+-----+-----+-----+ 1560  
 K K E P M A P S H S G I P K P G M K S M -  
 GCCCGGGAATCCCCAAGTGCCCCAGCGCTTCCAAGGAAGGGGAGCGGAGCCGGAGTGG  
 1561 -----+-----+-----+-----+-----+-----+-----+ 1620  
 P G K S P S A P A P S K E G E R S R S G -  
 GAAGCTGAGCTCAGGACTCCCCCAGCAGAAGCCCCAGCTGGACGGCAGACACTCCAGTTC  
 1621 -----+-----+-----+-----+-----+-----+-----+ 1680  
 K L S S G L P Q Q K P Q L D G R H S S S -  
 CTCTTCCAGCCTGGCGTCCTCAGAAGGAAAAGGCCCAGGAGGGACCACCCTGAACCACAG  
 1681 -----+-----+-----+-----+-----+-----+-----+ 1740  
 S S S L A S S E G K G P G G T T L N H S -

Fig. 17 (cont'd 1)

88/124

CATCAGCAGCCAGACTGTCAGTGGGTCTGTCTGGGACCACCCAGACCACAGGAAGCAATAC  
 1741 -----+-----+-----+-----+-----+-----+-----+ 1800  
 I S S Q T V S G S V G T T Q T T G S N T -  
 CGTCAGTGTTCAGCTACCTCAGCCCCAGCAGCAATACAACCATCCCAACACTGCCACGGT  
 1801 -----+-----+-----+-----+-----+-----+-----+ 1860  
 V S V Q L P Q P Q Q Q Y N H P N T A T V -  
 TGCACCTTTCCTGTACAGGTCTCAGACGGACACTGAAGGGAATGTTACTGCCGAGTCAAG  
 1861 -----+-----+-----+-----+-----+-----+-----+ 1920  
 A P F L Y R S Q T D T E G N V T A E S S -  
 CTC AACAGGTGTGAGCGTGGAGCCCAGCCACTTCACCAAGACTGGACAGCCTGCTCTGGA  
 1921 -----+-----+-----+-----+-----+-----+-----+ 1980  
 S T G V S V E P S H F T K T G Q P A L E -  
 AGAACTCACTGGGGAAGATCCTGAGGCTCGGCGGCTGCGGACAGTGAAGAACATCGCTGA  
 1981 -----+-----+-----+-----+-----+-----+-----+ 2040  
 E L T G E D P E A R R L R T V K N I A D -  
 TCTGCGGCAGAAATTTGGAGGAAACCATGTCCAGTTTAAGGGGAACCTCAGGTTACACACAG  
 2041 -----+-----+-----+-----+-----+-----+-----+ 2100  
 L R Q N L E E T M S S L R G T Q V T H S -  
 CACATTGGA AACCACGTTTGGACACCAATGTCAACCACGGAGATGAGTGGCCGTAGCATACT  
 2101 -----+-----+-----+-----+-----+-----+-----+ 2160  
 T L E T T F D T N V T T E M S G R S I L -  
 CAGCTTGACAGGGAGGCCCCACACCTCTGTCTGGAGACTGGGCCAGTCCAGCCCTCGGCT  
 2161 -----+-----+-----+-----+-----+-----+-----+ 2220  
 S L T G R P T P L S W R L G Q S S P R L -  
 CCAAGCAGGAGACGCCCCCTCAATGGGCAATGGGTATCCCCCTCGAGCCAACGCCAGCAG  
 2221 -----+-----+-----+-----+-----+-----+-----+ 2280  
 Q A G D A P S M G N G Y P P R A N A S R -  
 GTTCATCAACACTGAGTCAGGTCGCTATGTGTACTCCGCCCCCTCTGAGAAGGCAGCTGGC  
 2281 -----+-----+-----+-----+-----+-----+-----+ 2340  
 F I N T E S G R Y V Y S A P L R R Q L A -  
 CTCCCGGGCAGTAGTGTCTGCCACGTGGACGTCTCAGACAAGGCAGGAGATGAGATGGA  
 2341 -----+-----+-----+-----+-----+-----+-----+ 2400  
 S R G S S V C H V D V S D K A G D E M D -  
 CCTGGAAGGCATCAGCATGGACGCCCCCGGCTACATGAGCGATGGGGATGTTCTGAGCAA  
 2401 -----+-----+-----+-----+-----+-----+-----+ 2460  
 L E G I S M D A P G Y M S D G D V L S K -  
 GAACATCCGGACCGATGACATTACAAGCGGATACATGACTGATGGTGGACTTGGCCTCTA  
 2461 -----+-----+-----+-----+-----+-----+-----+ 2520  
 N I R T D D I T S G Y M T D G G L G L Y -  
 TACCCGTCGCCTGAACCGGCTCCCTGATGGGATGGCTGTGGTACGGGAGACCCTGCAACG  
 2521 -----+-----+-----+-----+-----+-----+-----+ 2580  
 T R R L N R L P D G M A V V R E T L Q R -  
 AAATACCTCCCTGGGCCTCGGAGACGCTGACAGCTGGGACGACAGCAGCTCCGTCAGCAG  
 2581 -----+-----+-----+-----+-----+-----+-----+ 2640  
 N T S L G L G D A D S W D D S S S V S S -

Fig. 17 (cont'd 2)

CGGCATCAGCGACACCATAGACAACCTCAGCACTGATGACATCAACACCAGCTCCTCCAT  
 2641 -----+-----+-----+-----+-----+-----+-----+ 2700  
 G I S D T I D N L S T D D I N T S S S I -  
 CAGCTCTTATGCCAACACACCTGCCTCCTCTCGAAAAACCTGGATGTGCAGACTGATGC  
 2701 -----+-----+-----+-----+-----+-----+-----+ 2760  
 S S Y A N T P A S S R K N L D V Q T D A -  
 TGAGAAGCACTCACAGGTGGAGAGGAATTCCTGTGGTCTGGTGATGATGTCAAGAAATC  
 2761 -----+-----+-----+-----+-----+-----+-----+ 2820  
 E K H S Q V E R N S L W S G D D V K K S -  
 AGACGGAGGCTCAGACAGCGGCATAAAAAATGGAGCCAGGTCCAAGTGGAGGCGGAATCC  
 2821 -----+-----+-----+-----+-----+-----+-----+ 2880  
 D G G S D S G I K M E P G S K W R R N P -  
 TTCTGATGTGTCTGACGAGTCCGACAAAAGCACGTCGGGCAAGAAGAATCCTGTTCATCTC  
 2881 -----+-----+-----+-----+-----+-----+-----+ 2940  
 S D V S D E S D K S T S G K K N P V I S -  
 CCAGACAGGCTCATGGCGGCGAGGCATGACAGCTCAGGTGGGCATCACCATGCCAAGGAC  
 2941 -----+-----+-----+-----+-----+-----+-----+ 3000  
 Q T G S W R R G M T A Q V G I T M P R T -  
 GAAGGCTTCAGCCCCGGCAGGCGCACTGAAGACCCCAGGAAGTGGAAAAACAGACGACGC  
 3001 -----+-----+-----+-----+-----+-----+-----+ 3060  
 K A S A P A G A L K T P G T G K T D D A -  
 AAAGGTGTCTGAGAAAGGAAGGCTTTCTCCTAAAGCCTCCCAGGTGAAGCGCTCCCCATC  
 3061 -----+-----+-----+-----+-----+-----+-----+ 3120  
 K V S E K G R L S P K A S Q V K R S P S -  
 AGATGCAGGCCGGAGCAGTGGTGACGAATCCAAAAAGCCCCTCCCCAGCAGCTCTAGGAC  
 3121 -----+-----+-----+-----+-----+-----+-----+ 3180  
 D A G R S S G D E S K K P L P S S S R T -  
 ACCTACTGCCAATGCCAACAGCTTTGGGTTCAAGAAGCAGAGTGGTTCCGCCACCGGCCT  
 3181 -----+-----+-----+-----+-----+-----+-----+ 3240  
 P T A N A N S F G F K K Q S G S A T G L -  
 GGCCATGATCACAGCCAGCGGGGTGACTGTCACCAGCAGGTCAGCCACACTGGGCAAAAT  
 3241 -----+-----+-----+-----+-----+-----+-----+ 3300  
 A M I T A S G V T V T S R S A T L G K I -  
 CCCAAAGTCATCTGCACTCGTCAGTCGGTCTGCTGGTCCGAAGTCAAGTATGGATGGGGC  
 3301 -----+-----+-----+-----+-----+-----+-----+ 3360  
 P K S S A L V S R S A G R K S S M D G A -  
 TCAGAATCAGGATGACGGGTATCTAGCCCTAAGCTCCCGGACAAACCTTCAGTACCGGAG  
 3361 -----+-----+-----+-----+-----+-----+-----+ 3420  
 Q N Q D D G Y L A L S S R T N L Q Y R S -  
 TTTGCCGAGGCCCAGTAAGTCCAACAGCCGAACGGGGCTGGGAACAGGTCTAGCACCAG  
 3421 -----+-----+-----+-----+-----+-----+-----+ 3480  
 L P R P S K S N S R N G A G N R S S T S -  
 CAGCATAGATTCCAACATTAGCAGCAAGTCCGCAGGCCTGCCAGTGCCCAAACTGAGGGA  
 3481 -----+-----+-----+-----+-----+-----+-----+ 3540  
 S I D S N I S S K S A G L P V P K L R E -

Fig. 17 (cont'd 3)

```

GCCTTCCAAAACAGCCCTAGGCAGCTCTCTACCAGGTCTGGTCAACCAAACAGACAAGGA
3541 -----+-----+-----+-----+-----+-----+-----+ 3600
      P S K T A L G S S L P G L V N Q T D K E -

GAAAGGCATCTCATCAGACAACGAGAGTGTGGCTTCCTGTAACCTCGGTGAAAGTGAATCC
3601 -----+-----+-----+-----+-----+-----+-----+ 3660
      K G I S S D N E S V A S C N S V K V N P -

GGCAGCCCAGCCTGTGTCCAGTCCGGCTCAGACCAGTCTCCAGCCTGGAGCCAAGTACCC
3661 -----+-----+-----+-----+-----+-----+-----+ 3720
      A A Q P V S S P A Q T S L Q P G A K Y P -

AGATGTGGCCTCTCCCACACTCCGCAGACTCTTTGGTGGGAAGCCTACCAAGCAAGTGCC
3721 -----+-----+-----+-----+-----+-----+-----+ 3780
      D V A S P T L R R L F G G K P T K Q V P -

CATCGCCACAGCTGAAAACATGAAAAATTCCGGTGGTCATCTCCAATCCTCATGCCACCAT
3781 -----+-----+-----+-----+-----+-----+-----+ 3840
      I A T A E N M K N S V V I S N P H A T M -

GACTCAGCAAGGTAACCTAGACTCCCCGTCAGGCAGTGGCGTCCTGAGCAGTGGGAGCAG
3841 -----+-----+-----+-----+-----+-----+-----+ 3900
      T Q Q G N L D S P S G S G V L S S G S S -

CAGTCCTCTCTACAGCAAGAATGTGGACCTCAACCAGTCTCCGCTAGCCTCCAGCCCCAG
3901 -----+-----+-----+-----+-----+-----+-----+ 3960
      S P L Y S K N V D L N Q S P L A S S P S -

CTCAGCCCAGTCCGCCCCCTTCCAACAGCCTCACCTGGGGCACCAACGCCAGCAGCTCCTC
3961 -----+-----+-----+-----+-----+-----+-----+ 4020
      S A H S A P S N S L T W G T N A S S S S -

CGCAGTTAGCAAGGATGGCCTGGGCTTTTCAGTCTGTCAGCAGCCTCCACACCAGCTGTGA
4021 -----+-----+-----+-----+-----+-----+-----+ 4080
      A V S K D G L G F Q S V S S L H T S C E -

GTCCATCGACATCTCCCTCAGCAGTGGAGGGGTCCCCAGCCACAATTCTTCCACTGGCCT
4081 -----+-----+-----+-----+-----+-----+-----+ 4140
      S I D I S L S S G G V P S H N S S T G L -

CATCGCCTCCTCCAAGGACGACTCCTTGACTCCCTTTGTCAGAACTAACAGTGTGAAGAC
4141 -----+-----+-----+-----+-----+-----+-----+ 4200
      I A S S K D D S L T P F V R T N S V K T -

CACACTGTCAGAAAGCCCTCTCTCTTCCCCTGCTGCTAGCCCTAAGTTCTGCAGAAGTAC
4201 -----+-----+-----+-----+-----+-----+-----+ 4260
      T L S E S P L S S P A A S P K F C R S T -

TCTGCCCAGGAAACAGGACAGTGACCCGCACCTTGATAGGAACACTTTGCCTAAGAAAGG
4261 -----+-----+-----+-----+-----+-----+-----+ 4320
      L P R K Q D S D P H L D R N T L P K K G -

ACTCAGGTATACTCCCACCTCCCAGCTTCGCACGCAAGAAGATGCAAAGAATGGTTACG
4321 -----+-----+-----+-----+-----+-----+-----+ 4380
      L R Y T P T S Q L R T Q E D A K E W L R -

GTCCCATTTCTGCAGGAGGCCTTCAGGACACCGCTGCCAATTCCCCCTTTTCTCTGGCTC
4381 -----+-----+-----+-----+-----+-----+-----+ 4440
      S H S A G G L Q D T A A N S P F S S G S -

```

Fig. 17 (cont'd 4)

91/124

```

CAGCGTGACTTCTCCCTCCGGAACAAGATTCAACTTTTCCAGCTTGCGAGTCCCACCAC
4441 -----+-----+-----+-----+-----+-----+ 4500
      S V T S P S G T R F N F S Q L A S P T T -

TGTCACCCAGATGAGCTTGTCCAACCCGACCATGCTGAGGACTCACAGCCTCTCCAATGC
4501 -----+-----+-----+-----+-----+-----+ 4560
      V T Q M S L S N P T M L R T H S L S N A -

TGATGGGCAGTATGATCCATACTGACAGCCGCTTCCGGAATAGCTCCATGTCCCTGGA
4561 -----+-----+-----+-----+-----+-----+ 4620
      D G Q Y D P Y T D S R F R N S S M S L D -

TGAGAAGAGCAGAACCATGAGCCGTTCCAGGCTCATTCCGGGATGGGTTTGAAGAAGTTCA
4621 -----+-----+-----+-----+-----+-----+ 4680
      E K S R T M S R S G S F R D G F E E V H -

TGGATCCTCACTCTCCTTGGTTTCCAGCACATCGTCAGTTTATTCTACACCAGAAGAAAA
4681 -----+-----+-----+-----+-----+-----+ 4740
      G S S L S L V S S T S S V Y S T P E E K -

ATGCCAGTCAGAGATTCGCAAGCTGCGGCGGGAAGTGGATGCCTCCCAGGAGAAAGTTTC
4741 -----+-----+-----+-----+-----+-----+ 4800
      C Q S E I R K L R R E L D A S Q E K V S -

AGCTTTGACCACCCAGCTGACAGCAAATGCTCACCTTGTGGCTGCCTTTGAACAGAGTCT
4801 -----+-----+-----+-----+-----+-----+ 4860
      A L T T Q L T A N A H L V A A F E Q S L -

TGGTAACATGACAATCAGGCTCCAGAGTCTGACCATGACAGCTGAGCAGAAGGATTGAGA
4861 -----+-----+-----+-----+-----+-----+ 4920
      G N M T I R L Q S L T M T A E Q K D S E -

ACTGAATGAGTTAAGAAAAACCATTGAGCTGCTAAAGAAACAGAACGCAGCTGCCCAGGC
4921 -----+-----+-----+-----+-----+-----+ 4980
      L N E L R K T I E L L K K Q N A A A Q A -

TGCCATTAATGGAGTAATTAACACACCTGAGCTCAACTGCAAAGGAAACGGCACTGCCCA
4981 -----+-----+-----+-----+-----+-----+ 5040
      A I N G V I N T P E L N C K G N G T A Q -

GTCTGCAGACCTCCGCATCCGCAGGCAGCACTCCTCAGACAGCGTCTCCAGCATCAACAG
5041 -----+-----+-----+-----+-----+-----+ 5100
      S A D L R I R R Q H S S D S V S S I N S -

TGCCACCAGCCACTCCAGTGTGGGCAGCAACATAGAGAGTGACTCAAAGAAGAAGAAGAG
5101 -----+-----+-----+-----+-----+-----+ 5160
      A T S H S S V G S N I E S D S K K K K R -

GAAGAAGTGGGTCAATGAGTTACGCAGCTCCTTCAAGCAAGCTTTCGGGAAGAAGAAGTC
5161 -----+-----+-----+-----+-----+-----+ 5220
      K N W V N E L R S S F K Q A F G K K K S -

CCCAAAATCTGCGTCCTCTCATTGAGATATTGAGGAGATGACGGATTCTTCTTTGCCTTC
5221 -----+-----+-----+-----+-----+-----+ 5280
      P K S A S S H S D I E E M T D S S L P S -

CTCACCAAAGTTACCACACAATGGGTCCACAGGTTCCACCCCACTGCTGAGGAATTCTCA
5281 -----+-----+-----+-----+-----+-----+ 5340
      S P K L P H N G S T G S T P L L R N S H -

```

Fig. 17 (cont'd 5)

92/124

```

CTCCAACCTCTCTAATTTTCAGAATGCATGGATAGTGAAGCTGAGACCGTCATGCAGCTCCG
5341 -----+-----+-----+-----+-----+-----+ 5400
      S N S L I S E C M D S E A E T V M Q L R -

AAATGAGTTAAGAGACAAGGAGATGAAGCTGACAGATATCCGCTTAGAAGCTCTCAGTTC
5401 -----+-----+-----+-----+-----+-----+ 5460
      N E L R D K E M K L T D I R L E A L S S -

TGCCCACCAGCTGGACCAGCTCCGGGAGGCCATGAACAGGATGCAGAGTGAAATAGAGAA
5461 -----+-----+-----+-----+-----+-----+ 5520
      A H Q L D Q L R E A M N R M Q S E I E K -

GCTGAAAGCTGAGAATGATCGGCTGAAGTCAGAGTCTCAAGGCAGTGGCTGCAGCCGGGC
5521 -----+-----+-----+-----+-----+-----+ 5580
      L K A E N D R L K S E S Q G S G C S R A -

TCCTTCCAAGTGTCCATCTCTGCCTCCCCGAGGCAGTCCATGGGCCTCTCCCAGCACAG
5581 -----+-----+-----+-----+-----+-----+ 5640
      P S Q V S I S A S P R Q S M G L S Q H S -

CTTGAACCTCACTGAGTCAACCAGCCTGGACATGTTGCTGGATGACACTGGTGAATGCTC
5641 -----+-----+-----+-----+-----+-----+ 5700
      L N L T E S T S L D M L L D D T G E C S -

GGCTCGGAAGGAAGGAGGCAGGCATGTTAAGATAGTTGTCAGCTTTCAGGAGGAAATGAA
5701 -----+-----+-----+-----+-----+-----+ 5760
      A R K E G G R H V K I V V S F Q E E M K -

GTGGAAGGAGGATTCCAGACCACATCTCTTTCTTATTGGCTGCATTGGAGTTAGTGGCAA
5761 -----+-----+-----+-----+-----+-----+ 5820
      W K E D S R P H L F L I G C I G V S G K -

GACGAAGTGGGATGTGCTCGATGGGGTGGTTAGACGGCTGTTCAAAGAATACATCATTCA
5821 -----+-----+-----+-----+-----+-----+ 5880
      T K W D V L D G V V R R L F K E Y I I H -

TGTCGACCCAGTGAGTCAGCTAGGGCTGAATTCAGACAGCGTTCTTGGCTACAGCATTGG
5881 -----+-----+-----+-----+-----+-----+ 5940
      V D P V S Q L G L N S D S V L G Y S I G -

AGAAATCAAGCGCAGCAACACTTCCGAAACACCGGAGCTGCTTCCTTGTGGCTATCTGGT
5941 -----+-----+-----+-----+-----+-----+ 6000
      E I K R S N T S E T P E L L P C G Y L V -

TGGAGAGAACACGACCATCTCAGTGAAGTGTGAAAGGGCTCGCAGAAAAACAGCCTGGACTC
6001 -----+-----+-----+-----+-----+-----+ 6060
      G E N T T I S V T V K G L A E N S L D S -

ACTGGTGTGTTGAGTCCTTGATTCCCAAGCCCATCCTGCAGCGCTACGTCTCCCTCCTGAT
6061 -----+-----+-----+-----+-----+-----+ 6120
      L V F E S L I P K P I L Q R Y V S L L I -

AGAGCACCGTCGGATCATTCTCTGCCCCAGCGGCACTGGGAAAACCTACCTGGCCAA
6121 -----+-----+-----+-----+-----+-----+ 6180
      E H R R I I L S G P S G T G K T Y L A N -

CCGGCTGTCTGAGTATATAGTGCTTCGAGAGGGACGGGAGTTGACAGACGGGGTTATCGC
6181 -----+-----+-----+-----+-----+-----+ 6240
      R L S E Y I V L R E G R E L T D G V I A -

```

Fig. 17 (cont'd 6)

CACCTTTAACGTGGACCATAAGTCCAGCAAGGAATTGCGCCAGTACCTGTCCAACCTTGC  
6241 -----+-----+-----+-----+-----+-----+ 6300  
T F N V D H K S S K E L R Q Y L S N L A -  
TGACCAGTGAACAGTGAGAACAAATGCTGTGGACATGCCCCCTCGTCATCATCCTGGACAA  
6301 -----+-----+-----+-----+-----+-----+ 6360  
D Q C N S E N N A V D M P L V I I L D N -  
CCTACACCACGTGAGCTCTCTGGGCGAGATCTTCAATGGGCTGCTCAACTGCAAGTACCA  
6361 -----+-----+-----+-----+-----+-----+ 6420  
GGATGTGGTGCACCTCGAGAGACCCGCTCTAGAAGTTACCCGACGAGTTGACGTTTCATGGT  
L H H V S S L G E I F N G L L N C K Y H -  
CAAATGCCCTTACATAAATTGGCACAATGAACCAGGCTACCTCTTCGACTCCCAACCTGCA  
6421 -----+-----+-----+-----+-----+-----+ 6480  
GTTTACGGGAATGTATTAACCGTGTACTTGGTCCGATGGAGAAGCTGAGGGTTGGACGT  
K C P Y I I G T M N Q A T S S T P N L Q -  
GCTTCACCATAACTTCAGATGGGTGCTTTGTGCCAACACACGGAGCCTGTGAAGGGTTT  
6481 -----+-----+-----+-----+-----+-----+ 6540  
CGAAGTGGTATTGAAGTCTACCCACGAAACACGGTTGGTGTGCCTCGGACACTTCCCAAA  
L H H N F R W V L C A N H T E P V K G F -  
CCTTGGCCGATTCTGAGGAGGAAGCTCATGGAAACAGAGATCAGTGGGCGGGTGCACAA  
6541 -----+-----+-----+-----+-----+-----+ 6600  
GGAACCGGCTAAGGACTCCTCCTTCGAGTACCTTTGTCTCTAGTCACCCGCCCACGCGTT  
L G R F L R R K L M E T E I S G R V R N -  
TATGGAGCTGGTAAAAATCATTGACTGGATTCCCAAGGTCTGGCATCACCTCAACCGCTT  
6601 -----+-----+-----+-----+-----+-----+ 6660  
ATACCTCGACCATTTTTAGTAAGTACCTAAGGGTTCCAGACCGTAGTGGAGTTGGCGAA  
M E L V K I I D W I P K V W H H L N R F -  
CCTGGAGGCTCACAGTTCCTCGGACGTCACCATCGGCCCCCGGCTCTTCCTGTCATGCCC  
6661 -----+-----+-----+-----+-----+-----+ 6720  
L E A H S S S D V T I G P R L F L S C P -  
CATCGATGTGGACGGCTCGAGAGTGTGGTTACCGACTTGTGGAACTATTCCATTATCCC  
6721 -----+-----+-----+-----+-----+-----+ 6780  
I D V D G S R V W F T D L W N Y S I I P -  
CTATCTCCTGGAAGCCGTCAGAGAAGGACTCCAGCTCTATGGAAGGCGCGCCCCCTGGGA  
6781 -----+-----+-----+-----+-----+-----+ 6840  
Y L L E A V R E G L Q L Y G R R A P W E -  
GGATCCTGCCAAGTGGGTGATGGACACATATCCATGGGCAGCCAGCCACAAACAGCACGA  
6841 -----+-----+-----+-----+-----+-----+ 6900  
D P A K W V M D T Y P W A A S P Q Q H E -  
GTGGCCTCCCCTGCTGCAGTTACGGCCTGAGGATGTGGGCTTCGACGGCTACTCCATGCC  
6901 -----+-----+-----+-----+-----+-----+ 6960  
W P P L L Q L R P E D V G F D G Y S M P -  
TCGGGAGGGATCGACAAGCAAGCAGATGCCCCCAGTGATGCTGAAGGTGACCCGCTGAT  
6961 -----+-----+-----+-----+-----+-----+ 7020  
R E G S T S K Q M P P S D A E G D P L M -  
GAACATGCTGATGAGGCTGCAGGAGGCAGCCAACTACTCCAGCCCCCAGAGCTATGACAG  
7021 -----+-----+-----+-----+-----+-----+ 7080  
N M L M R L Q E A A N Y S S P Q S Y D S -

Fig. 17 (cont'd 7)

CGACTCCAACAGCAACAGCCATCACGATGACATCTTGGACTCCTCTTTGGAGTCCACTCT  
7081 -----+-----+-----+-----+-----+-----+-----+ 7140  
D S N S N S H H D D I L D S S L E S T L -  
GTGACAGGGGGCCCGGAGCCAGCGCCCTCCTCTTCTCCTCACCGCATTCCACCTGCATCC  
7141 -----+-----+-----+-----+-----+-----+-----+ 7200  
\*  
CCCACATCACCTGAAGATGACTTCCTGAGCCAGCCCCCAGCCACAGCCTTAGAGCTGCG  
7201 -----+-----+-----+-----+-----+-----+-----+ 7260  
GGAACACCGAGACCCCCCGTCCTTCAGCCTCGACCTGGGTGCAGGCATCCCGGGCCAGCT  
7261 -----+-----+-----+-----+-----+-----+-----+ 7320  
GCCTGCGGACCGCTTCCTTCACAGCGAGAAGTGCCTACCTTCTGTTGTACTTTAATTA  
7321 -----+-----+-----+-----+-----+-----+-----+ 7380  
TTGTTTTGCCTTGTTGCTGTGACCTCCCTAAGACACTGAAGATACTTCTCGGGAAAGGAT  
7381 -----+-----+-----+-----+-----+-----+-----+ 7440  
CATCGCCGTTGAAATGAAAAGAGAGACAGAGAGAGAAAAAAGAGAACCCACATGAA  
7441 -----+-----+-----+-----+-----+-----+-----+ 7500  
GCTCTGAAACCAAACAGCATCCTGCCATGAGCTTCCCAGAGACAGAAGAGACTGGAGCAA  
7501 -----+-----+-----+-----+-----+-----+-----+ 7560  
AGTCGGAAACACAGAGAAGCACGGCTTCCCCTCAGCACAGACCCTCCAGACTGGGTCTCA  
7561 -----+-----+-----+-----+-----+-----+-----+ 7620  
GAGCCGTGCCACCCACCCTCCACACAGCCGGCCACAGGGAGAACTGGTGCTAACCAGGG  
7621 -----+-----+-----+-----+-----+-----+-----+ 7680  
TGCTTGCTTTGGTCACGTTCAACGCACTACAGAGCTACGACACAGGGGAACCTTAGGAGC  
7681 -----+-----+-----+-----+-----+-----+-----+ 7740  
AAATAAACCGTGCTTTCATGTTTTTTAAAAA  
7741 -----+-----+-----+-----+-----+-----+-----+ 7783

Fig. 17 (cont'd 8)



95/124

**T3-cDNA sequence and T3 protein encoded  
therein (isoform 2)**

```
AGCAGGGAGAGGGGAGGGAGGTGTGCCGTCTCTTCTGCAAGGGCAGTGCCCCAGCCTCAGC
1  -----+-----+-----+-----+-----+-----+ 60
  S R E R G G S V P S L L Q G Q C P S L S -

CACACTTCTGATCTGCAGTCCAACAGACCTTTCTAGCATGCCAAAGAGAACCTGGGGGTG
61  -----+-----+-----+-----+-----+-----+ 120
  H T S D L Q S N R P F * H A K E N L G V -

CCAGGGGGTCCTCAGAGCTCACACTGCACTTGTGGCACCCACAGCGAGTAGCCATCCGTG
121 -----+-----+-----+-----+-----+-----+ 180
  P G G P Q S S H C T C G T H S E * P S V -

AGCCGAGGAAACTGTACACAGATCTACACAGACTGGGCCAATCATTACCTAGCCAAATCC
181 -----+-----+-----+-----+-----+-----+ 240
  S R G N C T Q I Y T D W A N H Y L A K S -

GGCCACAAGCGTCTCATCAAGGATCTCCAGCAAGATGTGACAGATGGCGTCCTCCTGGCC
241 -----+-----+-----+-----+-----+-----+ 300
  G H K R L I K D L Q Q D V T D G V L L A -

CAGATTATCCAGGTTGTGGCAAATGAAAAGATTGAAGACATCAATGGCTGTCCGAAGAAC
301 -----+-----+-----+-----+-----+-----+ 360
  Q I I Q V V A N E K I E D I N G C P K N -

AGATCCCAAATGATTGAAAACATAGATGCCTGCTTGAATTTCCTGGCAGCTAAGGGAATA
361 -----+-----+-----+-----+-----+-----+ 420
  R S Q M I E N I D A C L N F L A A K G I -

AACATCCAGGGGCTGTCTGCAGAAGAGATCAGGAATGGAAACCTCAAGGCCATTCTAGGC
421 -----+-----+-----+-----+-----+-----+ 480
  N I Q G L S A E E I R N G N L K A I L G -

CTCTTCTTCAGCCTCTCCCGATACAAGCAGCAGCAGCAGCAGCCCCAGAAGCAGCACCTC
481 -----+-----+-----+-----+-----+-----+ 540
  L F F S L S R Y K Q Q Q Q Q P Q K Q H L -

TCCTCACCTCTGCCGCCCCGCGTATCCCAGGTGGCCGGGGCCCCCTCCCAGTGCCAGGCT
541 -----+-----+-----+-----+-----+-----+ 600
  S S P L P P A V S Q V A G A P S Q C Q A -

GGCACCCCTCAGCAGCAGGTGCCAGTCACTCCCCAAGCCCCGTGCCAGCCTCACCAGCCA
601 -----+-----+-----+-----+-----+-----+ 660
  G T P Q Q Q V P V T P Q A P C Q P H Q P -
```

Fig. 18

96/124  
T3 murine cDNA

ATGAGAAGAGCCGAACAATGAGTCGGTCAGGCTCCTTCCGGGATGGGTTTGAGGAAGTTC  
1 -----+-----+-----+-----+-----+ 60  
E K S R T M S R S G S F R D G F E E V H -

ATGGATCCTCCCTGTCCTTGGTTTCCAGCACATCCTCCATCTACTCCACGCCAGAAGAAA  
61 -----+-----+-----+-----+-----+ 120  
G S S L S L V S S T S S I Y S T P E E K -

AATGCCAGTCAGAGATTCGAAAGCTGAGGCGAGAACTGGATGCCTCCCAGGAAAAGGTGT  
121 -----+-----+-----+-----+-----+ 180  
C Q S E I R K L R R E L D A S Q E K V S -

AATGCCAGTCAGAGATTCGAAAGCTGAGGCGAGACGTGGATGCCTCCCAGGAAAAGGTGT  
121 -----+-----+-----+-----+-----+ 180  
C Q S E I R K L R R D V D A S Q E K V S -

CTGCGCTGACTACCCAGCTGACTGCAAATGCTCACCTTGTGGCAGCCTTCGAGCAGAGTC  
181 -----+-----+-----+-----+-----+ 240  
A L T T Q L T A N A H L V A A F E Q S L -

TGGGAAACATGACCATCAGGCTACAGAGTTTAACTATGACCGCTGAGCAGAAGGATTCAG  
241 -----+-----+-----+-----+-----+ 300  
G N M T I R L Q S L T M T A E Q K D S E -

AACTGAACGAGTTAAGAAAAACCATCGAGCTGCTGAAGAAACAGAATGCAGCTGCCCAGG  
301 -----+-----+-----+-----+-----+ 360  
L N E L R K T I E L L K K Q N A A A Q A -

CTGCCATTAATGGAGTGATTAACACGCCAGAGCTCAACTGCAAAGGAAATGGCAGTGCCA  
361 -----+-----+-----+-----+-----+ 420  
A I N G V I N T P E L N C K G N G S A R -

GGCTACAGACCTACGCATCCGCAGCAACACTCCTCCGACAGTGTCTCCAGTATCAATAGC  
421 -----+-----+-----+-----+-----+ 480  
L Q T Y A S A A T L L R Q C L Q Y Q \* R -

GCCACCAGCCACTCAAGTGTG  
481 -----+-----+ 501  
H Q P L K C -

Fig. 19

ACCCTCTGTCATGGCGCTAT  
Y P L S W R Y

ACAAATCCGGAAGCTTCGTA  
Q I R K L R

GCTTCGAAGTTCCTTCAACA  
L R S S F N

T2

CTCCATCAAGTCCTCCACCTC  
S I K S S T

TCCATCAAGTCCTCCACCTCG  
S I K S S T S

CCATCAAGTCCTCCACCTCGT  
S I K S S T S

CATCAAGTCCTCCACCTCGTC  
I K S S T S

ATCAAGTCCTCCACCTCGTCC  
I K S S T S S

TCAAGTCCTCCACCTCGTCCT  
I K S S T S S

CAAGTCCTCCACCTCGTCCTC  
K S S T S S

AAGTCCTCCACCTCGTCCTCC  
K S S T S S S

AGTCCTCCACCTCGTCCTCCG  
K S S T S S S

GTCTCCACCTCGTCCTCCGT  
S S T S S S

TCCTCCACCTCGTCCTCCGTG  
S S T S S S V

CCTCCACCTCGTCCTCCGTGG  
S S T S S S V

CTCCACCTCGTCCTCCGTGGG  
S T S S S V

TCCACCTCGTCCTCCGTGGGC  
S T S S S V G

CCACCTCGTCCTCCGTGGGCA  
T S S S V G

CACCTCGTCCTCCGTGGGCAC  
T S S S V G

ACCTCGTCCTCCGTGGGCACT  
T S S S V G G

CCTCGTCCTCCGTGGGCACTG  
T S S S V G G

CTCGTCCTCCGTGGGCACTGA  
S S S V G G

TCGTCTCCGTGGGCACTGAT  
S S S V G G T

CGTCCTCCGTGGGCACTGATG  
S S S V G G T

T2

98/124

AGTTGGAGGTGGACCTGCTGG  
L E V D L L

GTTGGAGGTGGACCTGCTGGA  
L E V D L L

TTGGAGGTGGACCTGCTGGAA  
L E V D L L E

TGGAGGTGGACCTGCTGGAAG  
L E V D L L E

GGAGGTGGACCTGCTGGAAGC  
E V D L L E

GAGGTGGACCTGCTGGAAGCA  
E V D L L E A

AGGTGGACCTGCTGGAAGCAG  
E V D L L E A

GGTGGACCTGCTGGAAGCAGA  
V D L L E A

GTGGACCTGCTGGAAGCAGAG  
V D L L E A E

TGGACCTGCTGGAAGCAGAGA  
V D L L E A E

GGACCTGCTGGAAGCAGAGAA  
D L L E A E

GACCTGCTGGAAGCAGAGAAT  
D L L E A E N

ACCTGCTGGAAGCAGAGAATG  
D L L E A E N

CCTGCTGGAAGCAGAGAATGA  
L L E A E N

CTGCTGGAAGCAGAGAATGAC  
L L E A E N D

TGCTGGAAGCAGAGAATGACC  
L L E A E N D

GCTGGAAGCAGAGAATGACCG  
L E A E N D

CTGGAAGCAGAGAATGACCGA  
L E A E N D R

TGGAAGCAGAGAATGACCGAC  
L E A E N D R

GGAAGCAGAGAATGACCGACT  
E A E N D R

GAAGCAGAGAATGACCGACTG  
E A E N D R L

T2

ATGACACCCAACCATGGCTTT  
M T P N H G F

TGACACCCAACCATGGCTTTC  
M T P N H G F

GACACCCAACCATGGCTTTCA  
T P N H G F

ACACCCAACCATGGCTTTTAC  
T P N H G F H

CACCCAACCATGGCTTTTCACT  
T P N H G F H

ACCCAACCATGGCTTTTCACTT  
P N H G F H

CCCAACCATGGCTTTTCACTTG  
P N H G F H L

CCAACCATGGCTTTTCACTTGA  
P N H G F H L

CAACCATGGCTTTTCACTTGAG  
N H G F H L

AACCATGGCTTTTCACTTGAGC  
N H G F H L S

ACCATGGCTTTTCACTTGAGCT  
N H G F H L S

CCATGGCTTTTCACTTGAGCTT  
H G F H L S

CATGGCTTTTCACTTGAGCTTC  
H G F H L S F

ATGGCTTTTCACTTGAGCTTCA  
H G F H L S F

TGGCTTTTCACTTGAGCTTCAG  
G F H L S F

GGCTTTTCACTTGAGCTTCAGG  
G F H L S F R

GCTTTTCACTTGAGCTTCAGGA  
G F H L S F R

CTTTTCACTTGAGCTTCAGGAT  
F H L S F R

TTTCACTTGAGCTTCAGGATG  
F H L S F R M

TTCACTTGAGCTTCAGGATGT  
F H L S F R M

TCACTTGAGCTTCAGGATGTT  
H L S F R M

Fig. 20 (cont'd 1)

T2

TAAAAGGTAAAAATGAAAAAC  
AAAAGGTAAAAATGAAAAACA  
AAAGGTAAAAATGAAAACAA  
AAGGTAAAAATGAAAACAAA  
AGGTAAAAATGAAAACAAAA  
GGTAAAAATGAAAACAAAA  
GTAAAAATGAAAACAAAAAC  
TAAAAATGAAAACAAAAACA  
AAAAATGAAAACAAAAACAA  
AAAAATGAAAACAAAAACAAG  
AAATGAAAACAAAAACAAGC  
AATGAAAACAAAAACAAGCA  
ATGAAAACAAAAACAAGCAA  
TGAAAACAAAAACAAGCAA  
GAAAACAAAAACAAGCAAAC  
AAAAACAAAAACAAGCAAACA  
AAACAAAAACAAGCAAACAA  
AAACAAAAACAAGCAAACAA  
AACAAAAACAAGCAAACAAAC  
ACAAAAACAAGCAAACAAACA

99/124

T2

CTCTAATTTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
CCTCTAATTTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
CTCTAATTTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
TCTAATTTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
CTAATTTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
TAATTTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
AATTTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
ATTTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
TTTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
TTAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
TAACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
AACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
ACAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
CAATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
AATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
ATGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
TGCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
GCATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
CATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
ATAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA  
TAAGAGTCAATAAACCCCTACTTTTTTAAAAAAA

Fig. 20 (cont'd 2)

T3

ACTGGGCCAATCATTACCTAG  
W A N H Y L  
TGCGGCCAATCATTACCTAGC  
W A N H Y L  
TGGGCCAATCATTACCTAGCC  
W A N H Y L A  
GGGCCAATCATTACCTAGCCA  
W A N H Y L A  
GGCCAATCATTACCTAGCCAA  
A N H Y L A  
GCCAATCATTACCTAGCCAAA  
A N H Y L A K  
CCAATCATTACCTAGCCAAAT  
A N H Y L A K  
CAATCATTACCTAGCCAAATC  
N H Y L A K  
AATCATTACCTAGCCAAATCC  
N H Y L A K S  
ATCATTACCTAGCCAAATCCG  
N H Y L A K S  
TCATTACCTAGCCAAATCCGG  
H Y L A K S  
CATTACCTAGCCAAATCCGGC  
H Y L A K S G  
ATTACCTAGCCAAATCCGGCC  
H Y L A K S G  
TTACCTAGCCAAATCCGGCCA  
Y L A K S G  
TACCTAGCCAAATCCGGCCAC  
Y L A K S G H  
ACCTAGCCAAATCCGGCCACA  
Y L A K S G H  
CCTAGCCAAATCCGGCCACAA  
L A K S G H  
CTAGCCAAATCCGGCCACAAG  
L A K S G H K  
TAGCCAAATCCGGCCACAAGC  
L A K S G H K  
AGCCAAATCCGGCCACAAGCG  
A K S G H K  
GCCAAATCCGGCCACAAGCGT  
A K S G H K R

T3

CGGCCACAAGCGTCTCATCAG  
G H K R L I  
GGCCACAAGCGTCTCATCAGG  
G H K R L I R  
GCCACAAGCGTCTCATCAGGG  
G H K R L I R  
CCACAAGCGTCTCATCAGGGA  
H K R L I R  
CACAAGCGTCTCATCAGGGAT  
H K R L I R D  
ACAAGCGTCTCATCAGGGATC  
H K R L I R D  
CAAGCGTCTCATCAGGGATCT  
K R L I R D  
AAGCGTCTCATCAGGGATCTC  
K R L I R D L  
AGCGTCTCATCAGGGATCTCC  
K R L I R D L  
GCGTCTCATCAGGGATCTCCA  
R L I R D L  
CGTCTCATCAGGGATCTCCAG  
R L I R D L Q  
GTCTCATCAGGGATCTCCAGC  
R L I R D L Q  
TCTCATCAGGGATCTCCAGCA  
L I R D L Q  
CTCATCAGGGATCTCCAGCAA  
L I R D L Q Q  
TCATCAGGGATCTCCAGCAAG  
L I R D L Q Q  
CATCAGGGATCTCCAGCAAGA  
I R D L Q Q  
ATCAGGGATCTCCAGCAAGAT  
I R D L Q Q D  
TCAGGGATCTCCAGCAAGATG  
I R D L Q Q D  
CAGGGATCTCCAGCAAGATGT  
R D L Q Q D  
AGGGATCTCCAGCAAGATGTG  
R D L Q Q D V  
GGGATCTCCAGCAAGATGTGA  
R D L Q Q D V

T3

CTGAAATGCAGTCCAGACTTC  
E M Q S R L  
TGAAATGCAGTCCAGACTTCC  
E M Q S R L  
GAAATGCAGTCCAGACTTCCA  
E M Q S R L P  
AAATGCAGTCCAGACTTCCAG  
E M Q S R L P  
AATGCAGTCCAGACTTCCAGG  
M Q S R L P  
ATGCAGTCCAGACTTCCAGGT  
M Q S R L P G  
TGCAGTCCAGACTTCCAGGTC  
M Q S R L P G  
GCAGTCCAGACTTCCAGGTCC  
Q S R L P G  
CAGTCCAGACTTCCAGGTCTT  
Q S R L P G P  
AGTCCAGACTTCCAGGTCTTAC  
Q S R L P G P  
GTCCAGACTTCCAGGTCTTACC  
S R L P G P  
TCCAGACTTCCAGGTCTTACC  
S R L P G P T  
CCAGACTTCCAGGTCTTACC  
S R L P G P T  
CAGACTTCCAGGTCTTACC  
R L P G P T  
AGACTTCCAGGTCTTACC  
R L P G P T A  
GACTTCCAGGTCTTACC  
R L P G P T A  
ACTTCCAGGTCTTACC  
L P G P T A  
CTTCCAGGTCTTACC  
L P G P T A R  
TTCCAGGTCTTACC  
L P G P T A R  
TCCAGGTCTTACC  
P G P T A R  
CCAGGTCTTACC  
P G P T A R V

Fig. 20 (cont'd 3)

T3

CGGGGCAGTAGTGTCTGCCAC  
R G S S V C H

GGGGCAGTAGTGTCTGCCACG  
R G S S V C H

GGGCAGTAGTGTCTGCCACGT  
G S S V C H

GGCAGTAGTGTCTGCCACGTG  
G S S V C H V

GCAGTAGTGTCTGCCACGTGG  
G S S V C H V

CAGTAGTGTCTGCCACGTGGA  
S S V C H V

AGTAGTGTCTGCCACGTGGAC  
S S V C H V D

GTAGTGTCTGCCACGTGGACG  
S S V C H V D

AGTGTCTGCCACGTGGACGT  
S V C H V D

AGTGTCTGCCACGTGGACGTC  
S V C H V D V

GTGTCTGCCACGTGGACGTCT  
S V C H V D V

TGTCTGCCACGTGGACGTCTC  
V C H V D V

GTCTGCCACGTGGACGTCTCA  
V C H V D V S

TCTGCCACGTGGACGTCTCAG  
V C H V D V S

CTGCCACGTGGACGTCTCAGA  
C H V D V S

TGCCACGTGGACGTCTCAGAC  
C H V D V S D

GCCACGTGGACGTCTCAGACA  
C H V D V S D

CCACGTGGACGTCTCAGACAA  
H V D V S D

CACGTGGACGTCTCAGACAAG  
H V D V S D K

ACGTGGACGTCTCAGACAAGG  
H V D V S D K

CGTGGACGTCTCAGACAAGGC  
V D V S D K

T3

101/124

TCACCATGCCAAGGACGAAGG  
T M P R T K

CACCATGCCAAGGACGAAGGC  
T M P R T K

ACCATGCCAAGGACGAAGGCT  
T M P R T K A

CCATGCCAAGGACGAAGGCTT  
T M P R T K A

CATGCCAAGGACGAAGGCTTC  
M P R T K A

ATGCCAAGGACGAAGGCTTCA  
M P R T K A S

TGCCAAGGACGAAGGCTTCAG  
M P R T K A S

GCCAAGGACGAAGGCTTCAGC  
P R T K A S

CCAAGGACGAAGGCTTCAGCC  
P R T K A S A

CAAGGACGAAGGCTTCAGCCC  
P R T K A S A

AAGGACGAAGGCTTCAGCCCC  
R T K A S A

AGGACGAAGGCTTCAGCCCCG  
R T K A S A P

GGACGAAGGCTTCAGCCCCGG  
R T K A S A P

GACGAAGGCTTCAGCCCCGGC  
T K A S A P

ACGAAGGCTTCAGCCCCGGCA  
T K A S A P A

CGAAGGCTTCAGCCCCGGCAG  
T K A S A P A

GAAGGCTTCAGCCCCGGCAGG  
K A S A P A

AAGGCTTCAGCCCCGGCAGGC  
K A S A P A G

AGGCTTCAGCCCCGGCAGGCG  
K A S A P A G

GGCTTCAGCCCCGGCAGGCGC  
A S A P A G

GCTTCAGCCCCGGCAGGCGCA  
A S A P A G A

T3

00314549 09/914549

AGAAGCAGAGTGGTTCGCCCA  
K Q S G S A

GAAGCAGAGTGGTTCGCCAC  
K Q S G S A

AAGCAGAGTGGTTCGCCACCC  
K Q S G S A T

AGCAGAGTGGTTCGCCACCCG  
K Q S G S A T

GCAGAGTGGTTCGCCACCCGG  
Q S G S A T

CAGAGTGGTTCGCCACCCGGC  
Q S G S A T G

AGAGTGGTTCGCCACCCGGCC  
Q S G S A T G

GAGTGGTTCGCCACCCGGCCT  
S G S A T G

AGTGGTTCGCCACCCGGCCTG  
S G S A T G L

GTGGTTCGCCACCCGGCCTGG  
S G S A T G L

TGGTTCGCCACCCGGCCTGGC  
G S A T G L

GGTTCGCCACCCGGCCTGGCC  
G S A T G L A

GTTCCGCCACCCGGCCTGGCCA  
G S A T G L A

TTCCGCCACCCGGCCTGGCCAT  
S A T G L A

TTCCGCCACCCGGCCTGGCCAT  
S A T G L A

TCCGCCACCCGGCCTGGCCATG  
S A T G L A M

CCGCCACCCGGCCTGGCCATGA  
S A T G L A M

CGCCACCCGGCCTGGCCATGAT  
A T G L A M

GCCACCCGGCCTGGCCATGATC  
A T G L A M I

CCACCCGGCCTGGCCATGATCA  
T G L A M I

CACCCGGCCTGGCCATGATCAC  
T G L A M I

ACCGGCCTGGCCATGATCACA  
T G L A M I T

T3

GGTCTGGTCAACCAACAGAC  
G L V N Q T D

GTCTGGTCAACCAACAGACA  
G L V N Q T D

TCTGGTCAACCAACAGACAA  
L V N Q T D

CTGGTCAACCAACAGACAAG  
L V N Q T D K

TGGTCAACCAACAGACAAGG  
L V N Q T D K

GGTCAACCAACAGACAAGGA  
V N Q T D K

GTGAACCAACAGACAAGGAG  
V N Q T D K E

TCAACCAACAGACAAGGAGA  
V N Q T D K E

CAACCAACAGACAAGGAGAA  
N Q T D K E

CCAACAGACAAGGAGAAAA  
N Q T D K E K

ACCAACAGACAAGGAGAAAG  
N Q T D K E K

CCAAACAGACAAGGAGAAAGG  
Q T D K E K

CAAACAGACAAGGAGAAAGGC  
Q T D K E K G

AAACAGACAAGGAGAAAGGCA  
Q T D K E K G

AACAGACAAGGAGAAAGGCAT  
T D K E K G

ACAGACAAGGAGAAAGGCATC  
T D K E K G I

CAGACAAGGAGAAAGGCATCT  
T D K E K G I

ACAAGGAGAAAGGCATCTC  
D K E K G I

GACAAGGAGAAAGGCATCTCA  
D K E K G I S

ACAAGGAGAAAGGCATCTCAT  
D K E K G I S

CAAGGAGAAAGGCATCTCATC  
K E K G I S

T3

102/124

TTTCATGGATCCTCACTCTCCT  
H G S S L S

TCATGGATCCTCACTCTCCTT  
H G S S L S

CATGGATCCTCACTCTCCTTG  
H G S S L S L

ATGGATCCTCACTCTCCTTGG  
H G S S L S L

TGGATCCTCACTCTCCTTGGT  
G S S L S L

GGATCCTCACTCTCCTTGGTT  
G S S L S L V

GATCCTCACTCTCCTTGGTTT  
G S S L S L V

ATCCTCACTCTCCTTGGTTTC  
S S L S L V

TCCTCACTCTCCTTGGTTTCC  
S S L S L V S

CCTCACTCTCCTTGGTTTCCA  
S S L S L V S

CTCACTCTCCTTGGTTTCCAG  
S L S L V S

TCACTCTCCTTGGTTTCCAGC  
S L S L V S S

CACTCTCCTTGGTTTCCAGCA  
S L S L V S S

ACTCTCCTTGGTTTCCAGCAC  
L S L V S S

CTCTCCTTGGTTTCCAGCACA  
L S L V S S T

TCTCCTTGGTTTCCAGCACAT  
L S L V S S T

CTCCTTGGTTTCCAGCACATC  
S L V S S T

TCCTTGGTTTCCAGCACATCG  
S L V S S T S

CCTTGGTTTCCAGCACATCGT  
S L V S S T S

CTTGGTTTCCAGCACATCGTC  
L V S S T S

TTGGTTTCCAGCACATCGTCA  
L V S S T S S

T3

09/914549

CTCCTTGGTTTCCAGCACATC  
S L V S S T

TCCTTGGTTTCCAGCACATCG  
S L V S S T S

CCTTGGTTTCCAGCACATCGT  
S L V S S T S

CTTGGTTTCCAGCACATCGTC  
L V S S T S

TTGGTTTCCAGCACATCGTCA  
L V S S T S S

TGGTTTCCAGCACATCGTCAG  
L V S S T S S

GGTTTCCAGCACATCGTCAGT  
V S S T S S

GTTTCCAGCACATCGTCAGTT  
V S S T S S V

TTTCCAGCACATCGTCAGTTT  
V S S T S S V

TTCCAGCACATCGTCAGTTTA  
S S T S S V

TCCAGCACATCGTCAGTTTAT  
S S T S S V Y

CCAGCACATCGTCAGTTTATT  
S S T S S V Y

CAGCACATCGTCAGTTTATTC  
S T S S V Y

AGCACATCGTCAGTTTATTCT  
S T S S V Y S

GCACATCGTCAGTTTATTCTA  
S T S S V Y S

CACATCGTCAGTTTATTCTAC  
T S S V Y S

ACATCGTCAGTTTATTCTACA  
T S S V Y S T

CATCGTCAGTTTATTCTACAC  
T S S V Y S T

ATCGTCAGTTTATTCTACACC  
S S V Y S T

TCGTAGTTTATTCTACACCA  
S S V Y S T P

CGTCAGTTTATTCTACACCAG  
S S V Y S T P

T3

GGAAGAAGTGGGTCAATGAGTTACGCAGCTCC  
K N W V N E L R S S

Fig. 20 (cont'd 5)



T3  
AAGAAGAAGAGGAAGAACTGGGTCAATGAGTTACGCAGCTCCTTCAAG  
K K K R K N W V N E L R S S F K

103/124

T3

TCTCTAATTTCAGAATGCATGGATA

T3

AGGAGATGAAGCTGACAGATATCCGCTTAGAAGCTCT

T3

GATTCCAGACCACACGTCTTTCTTATCG

Fig. 20 (cont'd 6)

104/124  
Alignment of the T protein family

09/914549

Note: The N-terminus of protein T2 was omitted in the alignment, since it has no significant homology to the T protein and the T3 protein.

```

T   ---MDLSSEMNRHGKNPVSHKLEDQK-----KIYTDWANHYLAKSGHKRLIKDLQ   60
T3  NQPERLNSQVLQGLQEPAGEGLPLRKSGSVENGFDQTQIYTDWANHYLAKSGHKRLIRDLO
T2  -----
      * * . . . * * . * . ***** . ***

T   QDIADGVLLAEIIQIIANEKVEDINGCPRSQSQMIENVDVCLSFLAARGVNVQGLSAEEI   120
T3  QDVTDGVLLAQIIQVVANEKIEDINGCCKNRSQMIENIDACLNFLAAKGINIQGLSAEEI
T2  -----
      ** . ***** . *** . ***** . ***** . * * * * * . * . *****

T   RGNLKAAILGLFFSLSRYKQQQ-HHQQQYYQS-----LVELQQRVT   180
T3  RGNLKAAILGLFFSLSRYKQQQQQPKQHLSSPLPPAVSQVAGAPSQCQAGTPQQQVPVT
T2  -----
      ***** . * . * * . * *

T   HASP-----PSEASQAKTQODMQSRLPGP-SRVPAAAGSSSKVQGASNLN--RRSQSFNSI   240
T3  PQAPCQPHQPAPHQQSKAQAEQMSRLPGPTARVSAAGSEAKTRGGSTTANNRRSQSFNNY
T2  -----
      . * * . * . * . * . ***** . * * * * * . * . * . *****

T   DKNK-----PP-----N-----   300
T3  DKSKPVTSPPPPPSSHEKEPLASSASSHPGMSDNAPASLESGSSSTPTNCSTSSAIPQPG
T2  -----
      ** * * *

T   -----N-----   360
T3  AATKPWRSKSLSVKHSATVSMLSVKPPGPEAPRPTPEAMKPAPNNQKSMLEKLKLFNSKG
T2  -----
      *

T   -----   420
T3  GSKAGEGPGSRDTSCERLETLPSEFEESEEELEASRMLTTVGPASSSPKIALKGIAQRTFS
T2  -----

T   --YAN-----GNEK-----   480
T3  RALTNNKKSSSLKGNEKEKEKQQREKDKESKDLAKRASVTERLDLKEEPKEDPSGAAPPEM
T2  -----
      . * * * *

T   -----   540
T3  PKKSSKIASFIPKGGKLNLSAKKEPMAPSHSGIPKPGMKSMGKSPSAPAPSKEGERSRSG
T2  -----

T   -----   600
T3  KLSSGLPQQKPQLDGRHSSSSSSSLASSEGKGPGGTTLNHSISSQTVSGSVGTTQTTGSNT
T2  -----

```

Fig. 21

T -----  
T3 VSVQLPQPQQQYNHPNTATVAPFLYRSQTDTEGNVTAESSSTGVSVEPSHFTKTGQPALE  
T2 -----

T --GEDPETRRMRTVKNIADLRQNLEETMSSLRGTOISHSTLETTFDSTVTTEVNGRTIP 720  
T3 ELTGEDPEARRLRTVKNIADLRQNLEETMSSLRGTOVTHSTLETTFDTNVTTEMSGRSIL  
T2 ----DPESQRKRTVQNVLDLRQNLEETMSSLRGSQVTHSSLEMTCYD--SDDANPRSVS  
\*\*\*\*\*.\* \*\*.\*. \*\*\*\*\*.\*..\*\*.\* \* . . \*

T NLTSRPTPMTWRLGQACPRLOAGDAPSLGAGYP-RSGTSRFIHTDPSRFMYTTPLRRAAV 780  
T3 SLTGRPTPLSWRLGQSSPRLQAGDAPSMGNGYPPRANASRFINTESGRYVYSAPLRRQLA  
T2 SLSNRSYPLSWRYGQSSPRLQAGDAPSVGGSCRSEGTPAWYMHGERAHYSHTMPMR--SP  
\* . \* \*..\*\* \*..\*\*\*\*\*.\* . . . . . \*

T SRLGNMSQIDMSEKA-SSDLDMSS-SEVDVGGYMSDGDILGKSLRTDDINSGYMTDGGGLNL 840  
T3 SRGSSVCHVDVSDKA-GDEMDLEGISMDAPGYMSDGDVLSKNIRTDITSGYMTDGGGLGL  
T2 SKLSHISRLELVESLDSDEVLDK-----SGYMSDSDLMGKMTEDDD----ITTG----  
\* . . . . . \*.. \*\*\*\*\* \*..\* . \*\* . \* \*

T YTRSLNRIPD-TATSRDIIQRGVHDTVVDADSWDDSSSVSSGLSDTLDNISTDDLNTTSS 900  
T3 YTRRLNRLPDGMAVVRETQQRNTSLGLGDADSWDDSSSVSSGISDTIDNLSTDDINTSSS  
T2 -----WDESSSISSGLSDASDNLSSEEFNASS  
\*\*.\*.\*.\*.\*.\*. \*\*.\*..\*..\*\*

T VSSYSNITVPSRKN--TQLRTDSEKRSTTDET--WDSP--EELKKPE--EDFDSHGDAG- 960  
T3 ISSYANTPASSRKN--LDVQTDAEKHSQVERNSLWSG--DDVKKSDGGSDSGIKMEPG-  
T2 LNSLPSTPTASRRNSTIVLRTDSEKRSLAESGLSWFSESEEKAPKKLEYDSGSLKMEPGT  
. \* \*\*.\* ..\*\*.\*.\* \* . \* . . \*

T GKWKTVSSGLPEDPEK-AGQKASLSVSQTSWRRGMSAQGGAPS--RQKAGTSALKTP- 1020  
T3 SKWRRNPDSVDESDEKSTSGKKNPVISQTSWRRGMTAQVGITMPRTKASAPAGALKTPG  
T2 SKWRRERPESCDDSSKGELKKPISLGHPPSLKKGKTPPVAVTSP--ITHTAQSALKVAG  
\*\* . . \* \* . . \*\*..\* . . \*\*\*

T -GKTDDAKASEKGKAPLKGSSQLQRSPSDAGKSSGDEGKK--PPSGIGRSTATSSFGFKKP 1080  
T3 TGKTDDAKVSEKGRLSPKASQVKRSPSDAGRSSGDESKKPLPSSSRTPANANSFGFKKQ  
T2 ---KPEGKATDKGKLAVKNTGLQRSSSDAGRDLSDAKK--PPSGIARPSTSGSFGYKKP  
. \* ..\*\* . \* ..\*\*.\*. . \*\* \* \* . . \*\*\*\*\*

T SG-VGSSAMITSSGATITSGSATLGKIPKSAAIGGKSNAGRKTSLDGSQONQDDVVLHVSS 1140  
T3 SGSATGLAMITASGVTVTSRSATLGKIPKSSALVSRS-AGRKSSMDGAQNQDDGYLALSS  
T2 PP-ATGTATVMQTG-----GSATLSKIQKSSGIPVKPVNGRKTSLDVSNSAEPGLAPGA  
\* . \* \*\*\*\*\* \* \* \* . . \*\*\*.\* \* . \*

T KTTLOQYRSLPRPSKSSTSGIPGR-GGHRSSSTSSID-SNVSSKSAGATTSKLREPTKIGSG 1200  
T3 RTNLQYRSLPRPSKSNR--NG--AGNRSSSTSSID-SNISSKSAGLPVPKLREPSKTALG  
T2 RSNIQYRSLPRPAKSSSMSVTGGRRGPRPVSSSIDPSLLSTKQGGLTSPRLKEPTKVASG  
.....\*\*\*\*\*.\* \* \* \* .\*\*\*\*\* \* . \* \* . \* . \*

T RSSPVTVNQTDKEKEKVAVSDSESVSLSG-SPKSSPTSASACG-AQGLRQPGSKYPDIAS 1260  
T3 SSLPGLVNQTDKEKG--ISSDNESVASCN-SVKVNPAAPVSSPAQTSLOPGAKYPDVAS  
T2 RTTPAPVNQTDREKE--KAKAKAVALDSDNISLKSIGSPESTPKNQASHPTATKLAELP  
. \* \*\*\*\*\*.\* . . \* . . \*

T PTFRRRLFAGAKAGGKSASAPNTEGVKSSSVMPSPSTTLARQGSLESPPSSGTGSMGSAGGLS 1320  
T3 PTLRRLFGGKP-TKQVPIATAENMKNSVVISNPHATMTQQGNLDSPLS-GSGVLS--S  
T2 PTPLRAT-AKSFVKPPSLANLDKVN-SNSLDLPSSSDTTHASKVPDLHATSSAS-----  
\*\* \* \* \* . . . \* . \* . . .

T GSSSPLFNKPSDLTTDVISLSHSLASSPASVHSFTSGGLVWAANMSSSSAGSKDTPSYQS 1380  
 T3 GSSSPLYSKNVDLN-----QSPLASSPSSAHSAPSNSLTWGTNASSSSAVSKDGLGFQS  
 T2 --GGPLPS-----CFTPSPAPILNINSASFSGLELMSGFSVPKETRMPK  
 \*\* \* \* \* \* \*  
 T MTSLHTSSESIDLPLS-----HHGSLSGLTTG-----THEVQSLLMRTGSRVSTLSES-- 1440  
 T3 VSSLHTSCESIDISLSSGGVPSHNSSTGLIASS-----KDDSLTPFVRTNSVKTTLSSESPL  
 T2 LSGLHRSMESLQMPMSLPSAFPSSTPVPTPPAPPAAPTEETEELTWGSPRAGQLDS--  
 .. \*\* \* \*\* . . \* . . \* . . \*  
 T -----MQLDRNTLPKKGLRYTPSSRQANQEEGKEWLRSHSTGGL 1500  
 T3 SSPAASPKFCRSTLPRKQDSDPHLDRNTLPKKGLRYTPTSQRLTQEDAKEWLRSHSAGGL  
 T2 -----NQRDRNTLPKKGLRYQLQS-----QEETKERRHSHTIGGL  
 . \*\*\*\*\* \* \*\* \* \* \*  
 T QDTGNQSPLVSPSAMSSSAAGKYHFSNLVSPTNLSQFNLPGPSMMRSNSIPAQDSSFDLY 1560  
 T3 QDTAANSFPSSGSSVTSPSGTRFNFSQLASPTTVTQMSLSNPTMLRTHSLSNADGQYDPY  
 T2 PESDDQSELPSPPALPMSLSAKGQLTNIVSPTAAT-----TPRITRSNSIPTHEAAFELY  
 .. \* \* . . . . . \* \* \* \* \*  
 T DDSQLCGSATSLEERPRAISHSGSFRDSMEEVHGSSLSLVSTSSLYSTAEKKAHSEQIH 1620  
 T3 TDSRFRNSSMSLDEKSRRTMSRSGSFRDGFEEVHGSSLSLVSTSSVYSTPEEKQSE-IR  
 T2 SGSQMG-STLSLAERPCKGMIRSGSFRDPTDDVHGSVLSLASSASSTYSSAEERMQSEQIR  
 \* . \* . \* \* . . . \*\*\*\*\* . \*\*\*\*\* \* \* \* \* \*  
 T KLRRELVASQEKVATLTSQLSANAHLVAAFEKSLGNMTGRLOSLTMTAEQKESELIELRE 1680  
 T3 KLRRELDASQEKVSALT'TQLTANAHLVAAFEQSLGNMTIRLOSLTMTAEQKDESELNELRK  
 T2 KLRRELESSQEKVATLTSQLSANANLVAAFEQSLVNMTSRLRHLAETAEEKDTELLDLRE  
 \*\*\*\*\* . \*\*\*\*\* . \* \* \* \* \* . \*\*\*\*\* \* \* \* \* \*  
 T TIEMLKAAQNSAAQAAIQGALNGPDHPPK-----DLRIRRHQHSSESVSINSATSHSS 1740  
 T3 TIBLLKKQNAQAQAAINGVINTPELNCKGNGTAQSADLRIRRHQSSDSVSSINSATSHSS  
 T2 TIDFLKKKNSEAAQAVIQGALNASETTPK-----ELRIKRONSSDSISSLNSITSHSS  
 \*\* . \* \* . \* \* \* \* \* . \* . \*\*\*\*\* \* \* \* \* \*  
 T IGSGNDADSKKKKKKNWL--RSSFKQAFGKKKSTKPPSSHSDIEELT--DSSLPASPKL 1800  
 T3 VGSNIESDSKKKKRKNWVNELRSSFKQAFGKKKSPKSASSHSDIEEMT--DSSLPSSPKL  
 T2 IGSSKDADAKKKKKKSWL--RSSFNKAFSIKKGPKSASSYSIDIEEIATPDSSAPSSPKL  
 . \* \* . . \* . \* \* \* \* . \* \* \* \* \*  
 T PHNAGDCGSASMKPSQSASAICTEA-----EAEIILQLKSELRE 1860  
 T3 PHNGSTGSTPLLRNHSNSLISECMDS-----EAETVMQLRNLERD  
 T2 QHGSTETASPSIKSSTSSSVGTDVTEGPAHPAPHTRLFHANEEEEPEKKEVSELSELWE  
 \* . . . \* \* \* . . . \* . \* \* \*  
 T KELKLTDIRLEALSSAHHLDOIREAMNRMQNEIEILKAENDRLKAETGNTAKPTRPPSES 1920  
 T3 KEMKLTDIRLEALSSAHQLDQLREAMNRMQSEIEKLKAENDRLKSES-QGSGCSRAPSQV  
 T2 KEMKLTDIRLEALNSAHQLDQLRETMHNMQLEVDLLEAENDRLKVAP--GPSSGSTPGQV  
 \*\* . \*\*\*\*\* \* \* \* \* \* . \* \* \* \* \*  
 T SSSTSSSSSRQSLGLSLNNLNITEAVSSDILLDDAGDATGHKDG-RSVKIIVSISKGYGR 1980  
 T3 SISAS--PRQSMGLSQHSLNLTESTSLDMLLDDTGECSARKEGGRHVKIVVSFQEEKMW  
 T2 PGSSALSSPRRSLGLALTHSFGPSLADTDLSPMDGISTCGPKEE-VTLRVVVRMPQHII  
 \* . . \* . \* \* \* . \* . \* . . . \*

T AKDQKSQAYLIGSIGVSGKTKWDVLDGVIRRLFKEYVFRIDTSTSLGLSSDCIASYCIGD 2040  
 T3 KEDSRPHLFLIGCIGVSGKTKWDVLDGVVRRRLFKEYIIHVDPVSQLGLNSDSVLGYSIGE  
 T2 KGD LKQQEFFLGCSKVSGKVDWKMLDEAVFQVFKDYISKMDPASTLGLSTESIHGYSISH  
 \* . . . . \* . . . . \* . . . . \* . . . . \* . . . . \*

T LIRSHNLEVPPELLPCGYLVGDNNIITVNLKGVEENSLDSFVFDTLIPKPITORYFNLLME 2100  
 T3 IKRSNTSETPELLPCGYLVGENTTISVTVKGLAENSLDSLVFESLIPKPILQRYVSLLE  
 T2 VKRVLDAEPPPEMPPCRRGVNN--ISVSLKGLKEKCVDSL VFETLIPKPMQHYISLLK  
 . \* . . . \* . . . . \* . . . . \* . . . . \* . . . . \*

T HHRIILSGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNVDHKSSKELOQYLANLAE 2160  
 T3 HHRIILSGPSGTGKTYLANRLSEYIVLREGRELTDGVIATFNVDHKSSKELRQYLSNLAD  
 T2 HRLVLSGPSGTGKTYLTNRLAEYLVERSGREVTEGIVSTFNMHQQCKDLQYLSNLAN  
 \* . . . . \* . . . . \* . . . . \* . . . . \* . . . . \*

T QCSADNNGVELPVVILDLNLHHVGSLSDFNGLNCKYNKCPYIIGTMNQGVSSSPNLEL 2220  
 T3 QCSENNAVDMPLVILDLNLHHVSSLGEIFNGLNCKYHKCPYIIGTMNQATSSTPNLQL  
 T2 QIDRETGIGDVPLVILLDDLSEAGSISELVNGALTCKYHKCPYIIGTTNQPVKMTPNHGF  
 \* . . . . \* . . . . \* . . . . \* . . . . \* . . . . \*

T HHNFRWVLCANHTEPVKGFGLGRYLRRKLIEIEIERNIRNNDLVKIIDWIPKTWHHLNSFL 2280  
 T3 HHNFRWVLCANHTEPVKGFGLGRFLRRKLMETEISGRVRNMELVKIIDWIPKVWHHLNRFL  
 T2 HLSFRMLTFSNNVEPANGFLVRYLRRKLVESDSDINANKEELLRLVDWVPKLWYHLHTFL  
 \* . . . . \* . . . . \* . . . . \* . . . . \* . . . . \*

T ETHSSSDVTIGPRLFLPCPMDVEGSRVWFMDLWNYSLVPIYLEAVREGLQMYGKRTPWED 2340  
 T3 EAHSSSDVTIGPRLFLSCPIDVDGSRVWFTDLWNYSIIPYLLEAVREGLQLYGRRAPWED  
 T2 EKHSTSDFLIGPCFFLSCPIGIEDFRTWFDLWNNSIIPYLQEGAKDGIKVHGQKAAWED  
 \* . . . . \* . . . . \* . . . . \* . . . . \* . . . . \*

T PSKWVLDTPWSSATLPQESPALLQLRPEDVGYESCTSTKEATTSKHIPQTDTEGDPLMN 2400  
 T3 PAKWVMDTPWAASPQQHEWPPLLQLRPEDVGFDGYSMPREGSTSKQMPPSDAEGDPLMN  
 T2 PVEWVRDTLPWPSAQDQSK--LYHLPPTVGPHSIASPPEDRTVKDSTPSSLDSDPLMA  
 \* . . . . \* . . . . \* . . . . \* . . . . \* . . . . \*

T MLMKLQEAANYSSSTQSCDSES--TSHHEDILDSSLESTL  
 T3 MLMRLQEAANYSSPQSYDSDSNSNSHHDDILDSSLESTL  
 T2 MLLKLQEAANYIE--SPDRET-----ILDPNLQATL  
 \* . . . . \* . . . . \* . . . . \* . . . . \*

Fig. 21 (cont'd 3)

## Alignment of the T protein with the POM121 protein

	10	20	30	40	50	60
T-Protein POM121	MDLSSEMNRHGKNPVSHKLEDQKKIYTDWANHYLAKSGHKRLIKDLQQDIADGVLLAEII					
T-Protein POM121	QIIANEKVEDINGCPRSQSQMIENVDVCLSFLAARGVNVQGLSAEEIRNGNLKAILGLFF					
	---MSPAAAAADGGERRRP-----PLGVREGRGR-TRGCGGPAGAAALGLALLGLAL					
	* * * * *					
T-Protein POM121	SLSRYKQQQHHQQYYQSLVELQQRVTHASPPSEASQAKTQODMQSRLPGPSRVPAAGSS					
	YLV-----P---AAAALAWLAVGASAAWWGLSREPRGP--					
	* * * *					
T-Protein M121	SKVQASNLNRRSQSFNSIDKNKPPNYANGNEKGEDPETRRMRTVKNIAIDLQNLEETMS					
	---RGLSSFVRESR-----RHPRPALASPLPAKSP-----VNGSLCBPRS					
	* * * * *					
T-Protein POM121	SLRGTOISHSTLETTDFDSTVTTEVNGRTIPNLTSRPTPMTWRLGQACPRLOAGDAPSLGA					
	PLGGPDPAELLMLGSGYL-----KPGPPEPALPQD-PRDRPGRPPPSRS					
	* * * * *					
T-Protein POM121	GYPRSGTSRFIHTDPSRFMYTTPLRRAAVSRLGNMSQIDMSEKASSDLMSSEVDVGGYM					
	PPSSSTAQRVHHVYP---ALPTPLLRPSRR-----PPHRDCGPLS					
	* * * * *					
T-Protein POM121	SDGDILGKSLRTDDINSGYMTDGGNLNLYTRSLNRIPDTATSRDIIQRGVHDVTVDADSWD					
	SRFVITPR-RRYPIQQAQYSLLGALPTVCWNGGHKKAVLSARNS-RMVCSPVTVRIAPPD					
	* * * * *					
Protein POM121	DSSSVSSGLSDTLNISTDDINTSSVSSYSNITVPSRKNTQLRTDSEKRSTTDETWDSP					
	-----SKLFRSPMPEQILSTTLSSPSSNAPDPCAKETVLNALKEKKKRTVAEEDQ-					
	* * * * *					
T-Protein POM121	EELKKPEEDFDSDHGDAGGKWKTVSSGLPEDPEKAGQKASLSVSQTGSWRRGMSAQQGAPS					
	LHLDGQENKRRRHDSG-----SGHSAFEPLVANGVPAAFVPKPGSLKRSLASQSSDDH					
	* * * * *					
T-Protein POM121	RQK-AGTSALKTPGKTDDAKASEKGK-APLKGSSLQSPSDAGKSSGDEGKKPPSGIGRS					
	LNKRSRTSSVSSLTSTCTGGIPSSSRNAITSSYSSTRGVSQWLKRSRG-PTSSPFSSPASS					
	* * * * *					
T-Protein POM121	TATSSFGFKKPSGVGSSAMITSSGATITSGSATLGKIPKSAAGGKSNAGRKTSLDGSQN					
	RSQTPERPAKKTREBEFCHQSSSSAPLVTDKESPGEKVTDPATGKQQLWTSPTTPGSSG					
	* * * * *					
T-Protein POM121	QDDVVLHVSSKTTTLQYRSLRPSKSSSTSGIPGRGGRSSTSSIDSNVSSKSAGATTSKLR					
	QRKRKIQLLPSRRGDQTLPPP-----P--ELG--YSITAEDLDMERR--AS--LQ					
	* * * * *					

Fig. 22

T-Protein  
POM121 EPTKIGSGRSSPVTVNQTDKEKEKVAVSDSESVLSGSPKSSPTSASACGAQGLRQPGSK  
WFKNVLEDKTDASTPATDTSP---ATSPFFTLTL---P---TVGPAASPASLPAPSS-  
\* . . . . \* \* . . . \* \* . . . \* \* . . . \*

T-Protein  
POM121 YPDIASPTFRRLFGAKAGGKSASAPNTEGVKSSSVMPSPSTTLARQGSLESPSSGTGSMG  
-----NPLLESCLKMQESPAPSSSEPPE--AATVAAPSPPKTPSLLAPLVSP-----  
\* \* . . . \* . . . \* \* . . . \* \* . . . \*

T-Protein  
POM121 SAGGLSGSSSPLFNKPSDLTTDVISLSHSLASSPASVHSFTSGGLVWAANMSSSSAGSKD  
----LTG---PLASTSSDSKPTTTFLGLASASSATPLTDTKAPGVSQAQLCVSTPAATAP  
\* . \* \* \* \* \* . . . \* . . . \* . . . \*

T-Protein  
POM121 TFSYQSMSTSLHTSSESIDLPLSHHGSLSGLTGTHEVQSLLMRTGSVRSTLSESMQLDRN  
SP-----TPASTLFGMLSPPASSSSSLATPGFACASPMFKPIFPATPKSE----SDN  
\* . . . . \* \* . . . \* . . . \* . . . \*

T-Protein  
POM121 TLPKKGLRYTPSSROANQEEGKEWLRSHSTGGLODTGNQSPLVSPSAMSSAAGKYHFSN  
PLP-----TSSSAATTTTASTALPTTATATAHTFKPIFESVEPFAAMP-----  
\* \* . . . \* . . . \* . . . \* . . . \*

T-Protein  
POM121 LVSPTNLSQFNLPGPSMMRSNSIPAQDSSFDLYDDSQLCGSATSLEERPRAISHSGSFRD  
LSPPFSLKQTTAPATTAATSAPLLTG-----L-----GTATST-----VATGTTAS  
\* \* \* \* . \* . . . \* . . . \* . . . \*

T-Protein  
POM121 SMEEVHGSSSLVLSSTSSLYSTAEKAHSEIHKLRRELVASQEKVATLTSQLSANAHLV  
ASKPVFGFVTTAASTASTIAS-----TSQSILFGGAPPVTASSAPALASIFQFGKPLA  
\* \* . . . \* . . . \* . . . \* . . . \*

T-Protein  
POM121 AAFEKSLGNMTGRLQSLTMTAEQKESELIELRETIEMLKQNSAAQAAIQGALNGPDHPP  
PAASVAGTSFSQSLASSAQTAASNSS--GGFSGFGGTLTTSTAPATTSQPTLTFSNTVT  
\* . . . . \* \* . . . \* . . . \* . . . \*

T-Protein  
POM121 KDLRIRRHSSSE-SVSSINSATSHSSIGSGNDADSKKKKKKNWLRSSFQAFGKKKSTK-  
PTFNIPFSASAKPALPTYPGANSQPTFG-ATDGATKP-----ALAPSFSSSTFGNSVAS  
\* . . . . \* . . . \* . . . \* . . . \*

T-Protein  
POM121 PPSSHSIDIEELTDSSLPASPKLPHNAGDCGSASMKPSQSASAICTEAEAEIILQLKSE  
APSAAPAPAAFGGAAQPAFGGLKASASTFG---TPASTQPAFGSTTS-----VFSFGSA  
\* . . . . \* . . . \* . . . \* . . . \*

T-Protein  
POM121 LREKELKLTDIRLEALSSAHHLDOIREAMNRMONEIBILKAENDRLKAETGNTAKPTRPP  
TTS-----GFGAAAATTQTHSGS-----SSSLFGSSTPS-PF  
\* . . . . \* . . . \* . . . \* . . . \*

T-Protein  
POM121 SESSSSTSSSSSRQSLGLSLNNLNITEAVSSDILLDDAGDATGHKDGSRVKIIVSISKGY  
TFGGSAAPAGGG---GFGLSATPGTGSTSGTFSFGSGQSGT---TGTTTSFGGSLSQNT  
\* . . . . \* . . . \* . . . \* . . . \*

T-Protein  
POM121 GRAKDQKSQAYLIGSIGVSGKTKWDVLDGVIRRLFKEYVFRIDTSTSLGLSSDCIASYCI  
LGAPSQSS--PFAFSVGSTPESKP-----VFGGTSTPTFGQSAPAPG---V  
\* \* \* \* . \* . . . \* . . . \* . . . \*

T-Protein  
POM121

GDLIRSHNLEVPPELLPCGYLVGDNNIITVNLKGVEENSIDSFVFDTLIPKPITQRYFNLL  
GTTGSSLSFGAPSTPAQGFVG-----VGPFPGSGAPSFSSIGAGSKTPGARQRLQAR  
\* \* \* \* \*

T-Protein  
POM121

MEHHRIILSGPSGTGKTYLANKLAEYVITKSGRKKTEDAIATFNVDHKSSKELQQYLANL  
RQHTRKK-----  
\* \*

T-Protein  
POM121

AEQCSADNNGVELPVVIIIDNLHHVGSLSDFNGFLNCKYNKCPYIIGTMNQGVSSSPNL  
-----

T-Protein  
POM121

ELHHNFRWVLCANHTEPVKGFLGRYLRRKLIEIEIERNIRNNDLVKIIDWIPKTWHHLNS  
-----

T-Protein  
POM121

FLETHSSSDVTIGPRLEFLPCPMDVEGSRVWFMDLWNYSLVPILEAVREGLQMYGKRTPW  
-----

T-Protein  
POM121

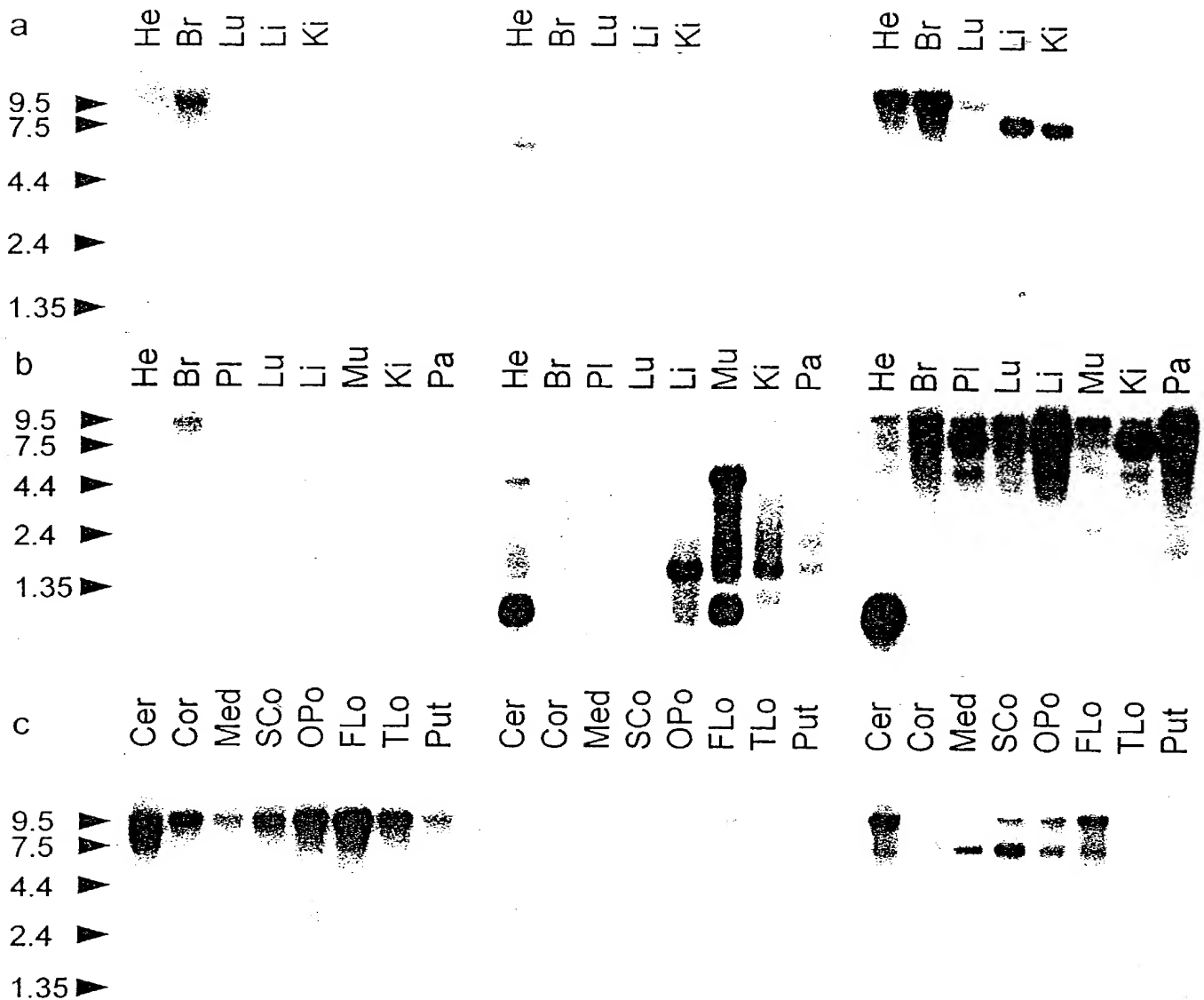
EDPSKWVLDTYPWSSATLPQESPALLQLRPEDVGYESCTSTKEATTSKHIPQTDTEGDPL  
-----

T-Protein  
POM121

MNMLMKLQEAANYSSSTQSCDSESTSHHEDILDSSLESTL  
-----

Fig. 22 (cont'd 2)





Expression of the T gene family.

**a** fetal tissue: left: T gene; middle: T2 gene; right: T3 gene.

He = heart; Br = brain; Lu = lungs; Li = liver; Ki = kidney

**b** adult tissue: left: T gene; middle: T2 gene; right: T3 gene.

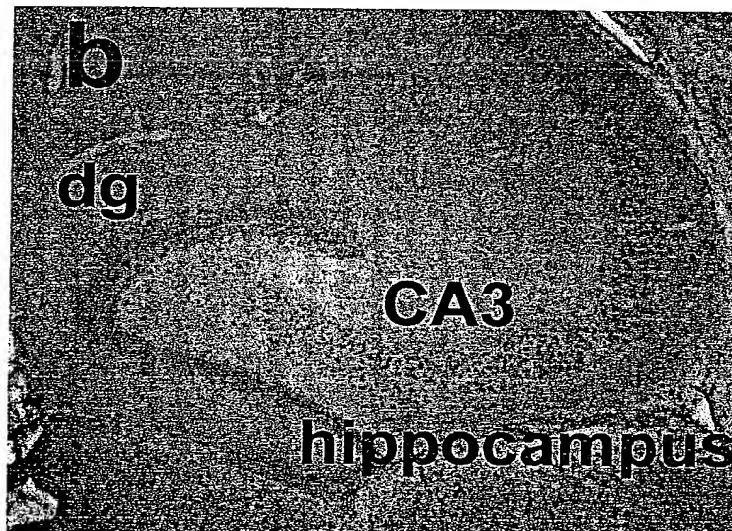
He = heart; Br = brain; Pl = placenta; Lu = lungs; Li = liver; Mu = skeletal muscle; Ki = kidney; Pa = pancreas

**c** adult brain regions: left: T gene; middle: T2 gene; right: T3 gene.

Cer = cerebellum; Cor = cerebral cortex; Med = medulla; Sco = spinal cord; Opo = occipital pole; Flo = frontal lobe; Tlo = temporal lobe; Put = putamen

Fig. 23

Fig. 24



113/124

Fig. 24

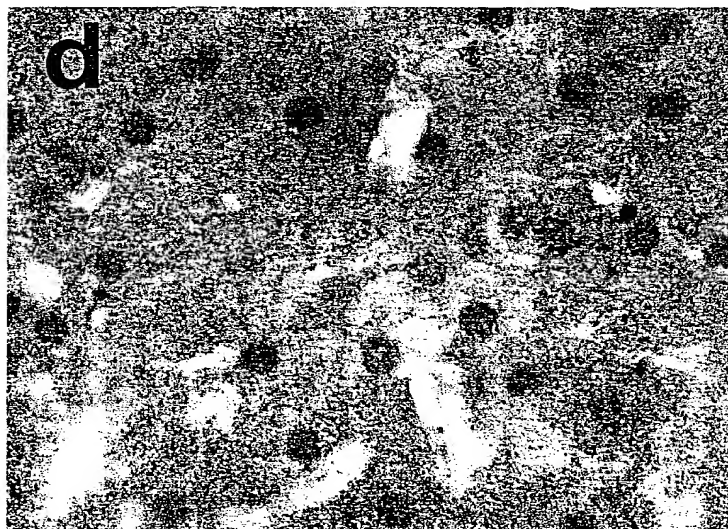
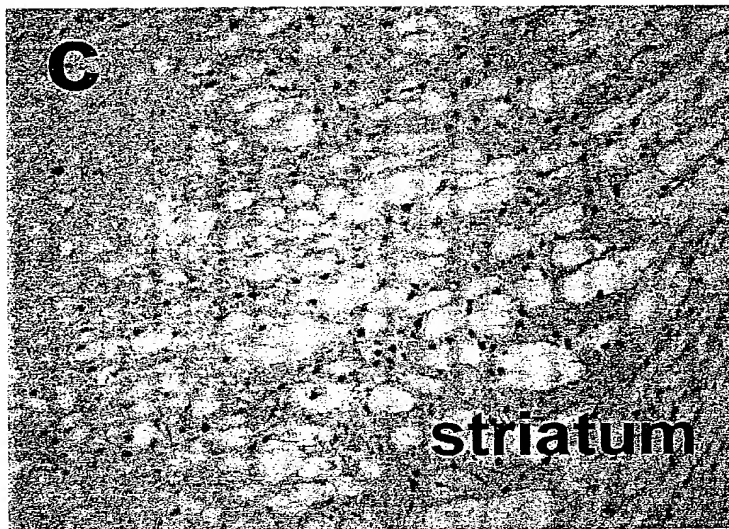


Fig. 24

114/124

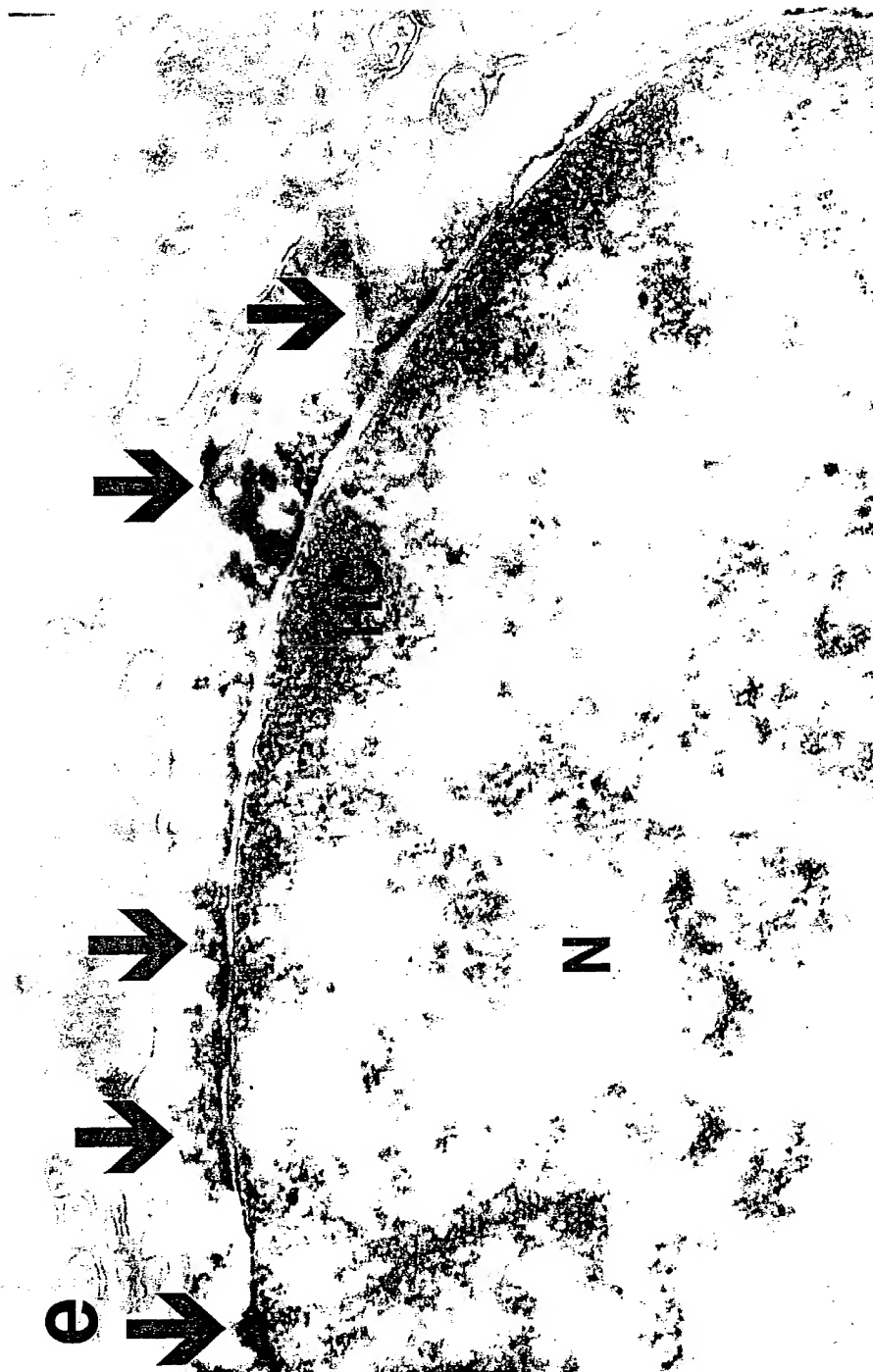


Figure legend of immunohisto and electron microscopy:

- a = brain stem. CG central grey = central grey of the brain stem
- b = hippocampus. dg = dental gyrus; CA3 cornu ammonis 3, both subregions of the hippocampus formation
- c = electronmicroscopic picture. N = nucleus, Hc heterochromatin

Fig. 25

115/124

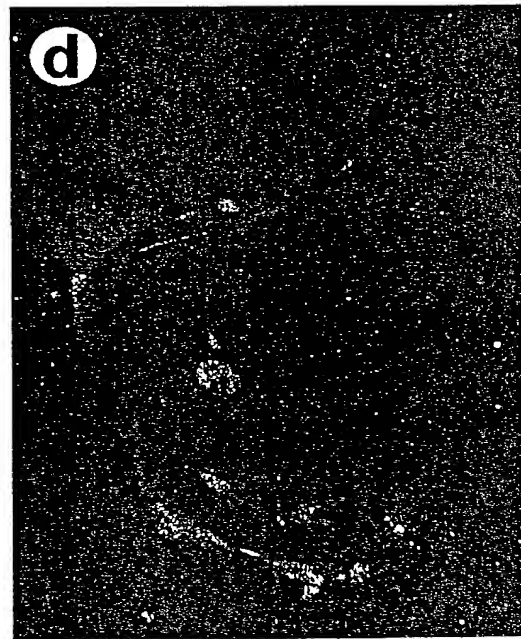
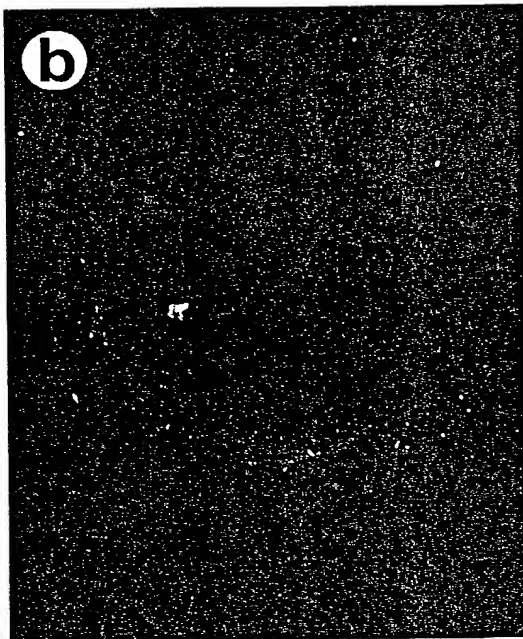
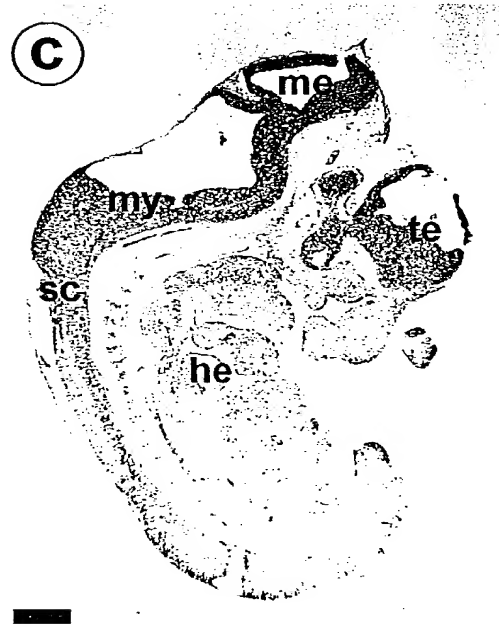
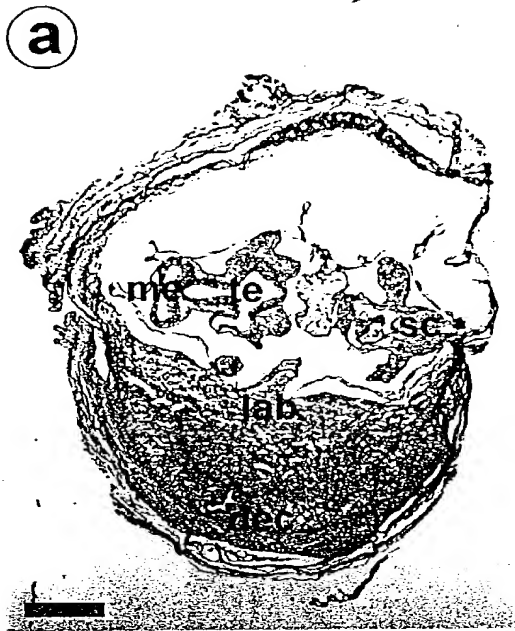


Fig. 25

116/124

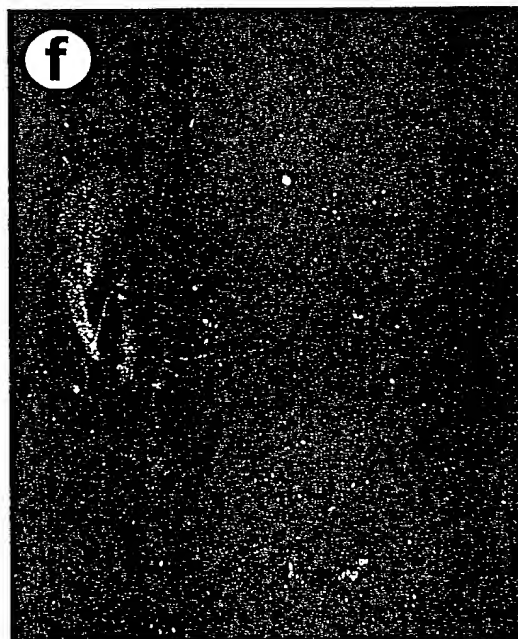
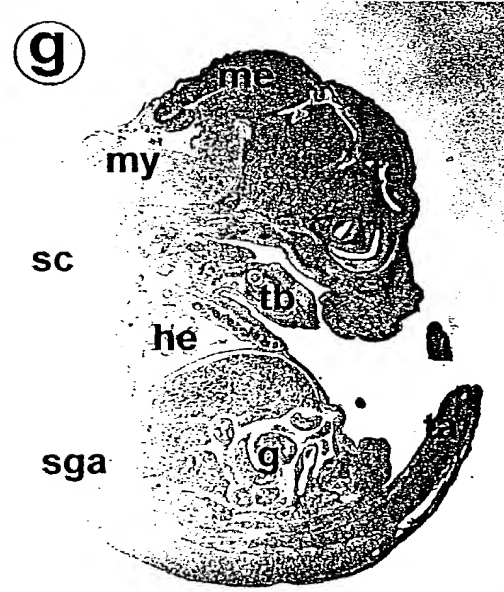
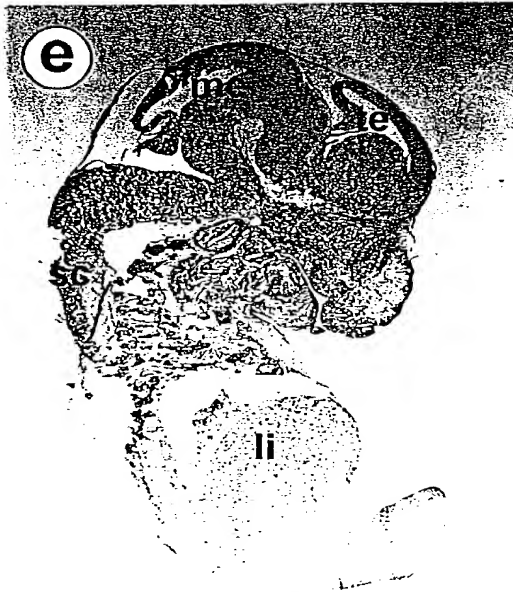


Fig. 26

117/124

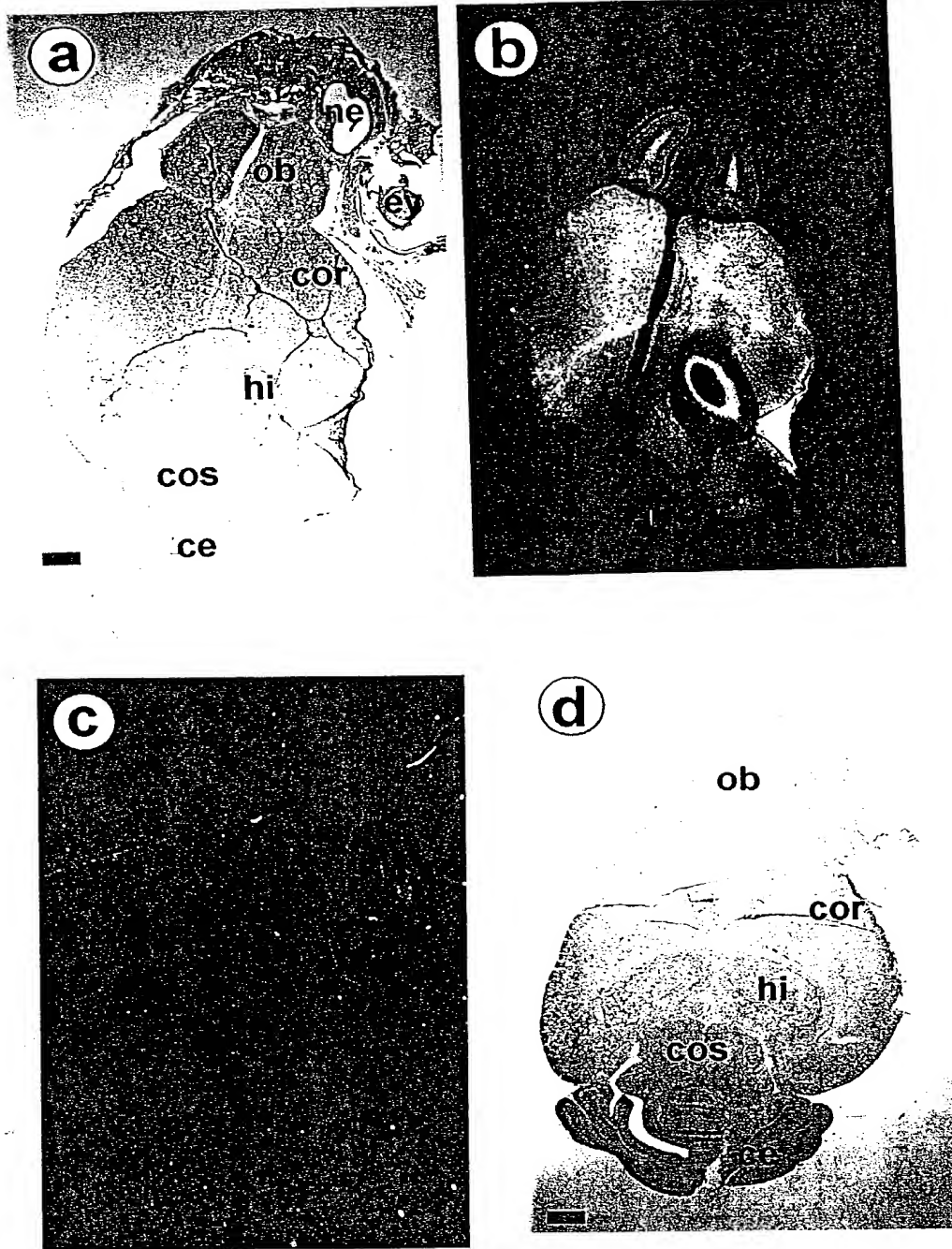




Fig. 26

118/124

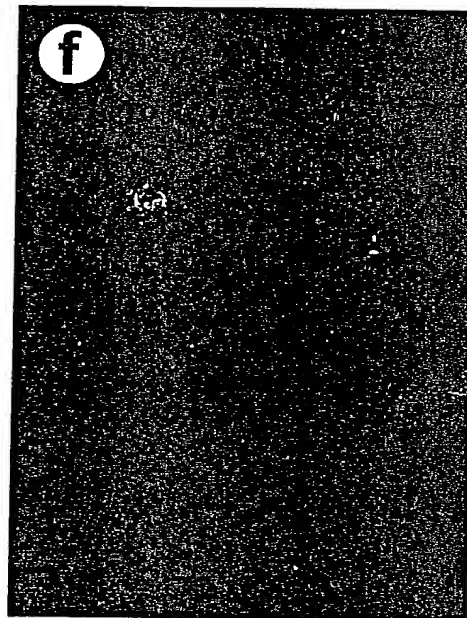




Fig. 27

119/124

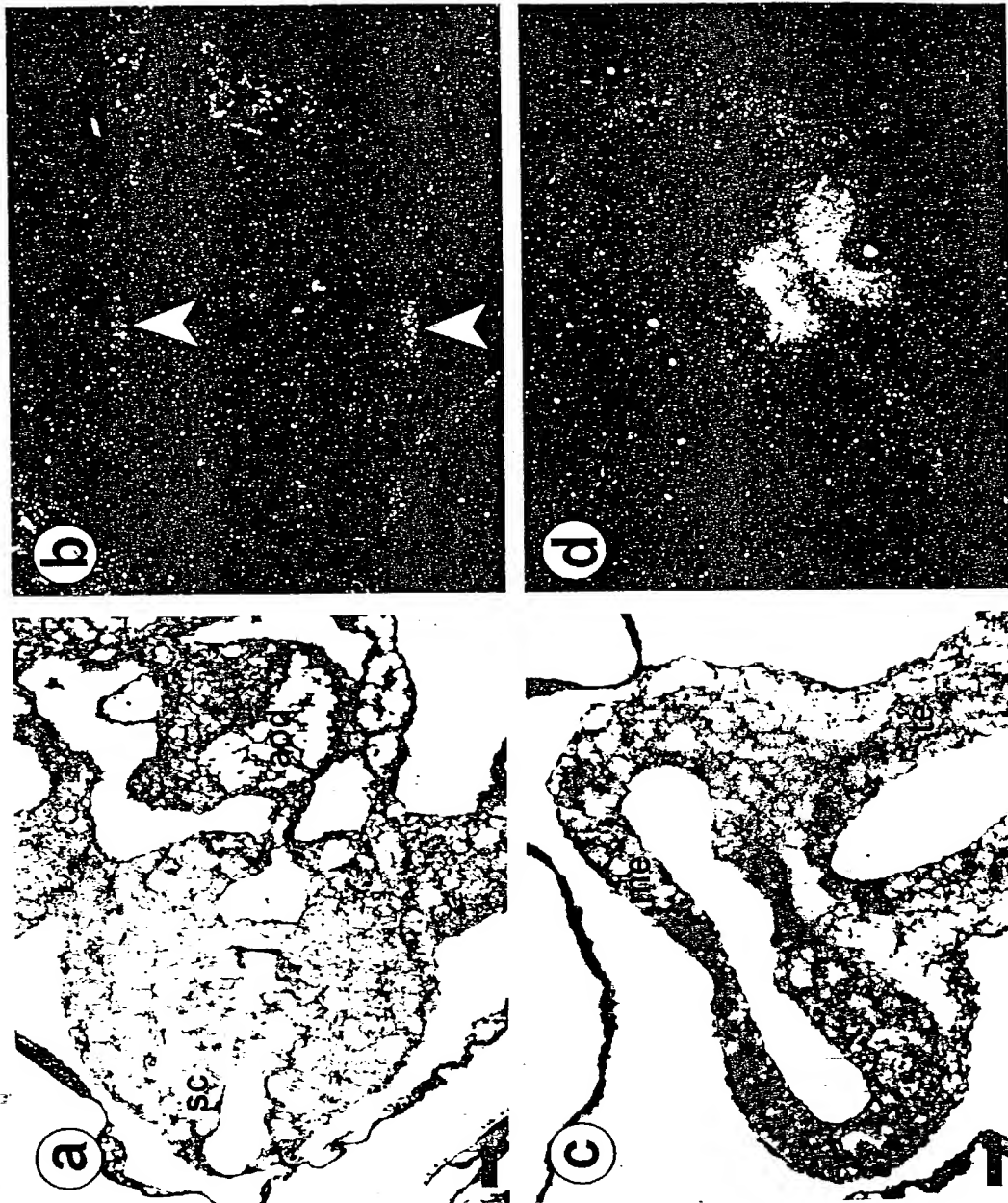


Fig. 27

120/124

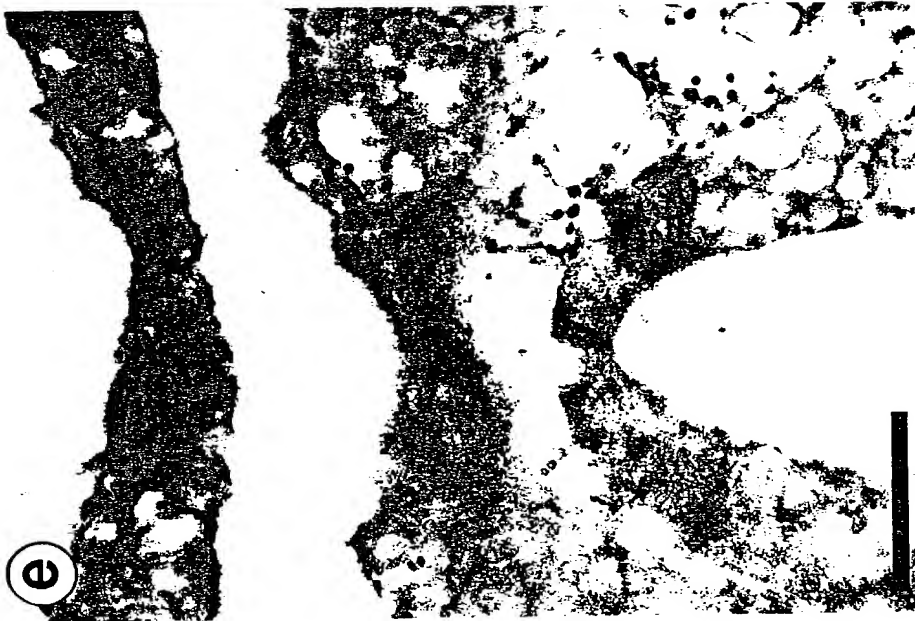
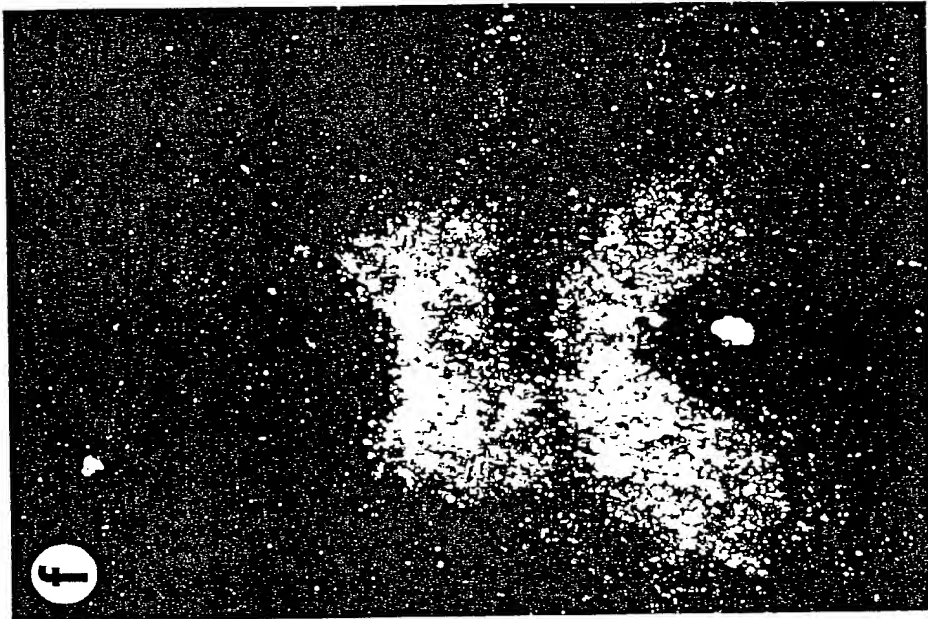


Fig. 28

121/124

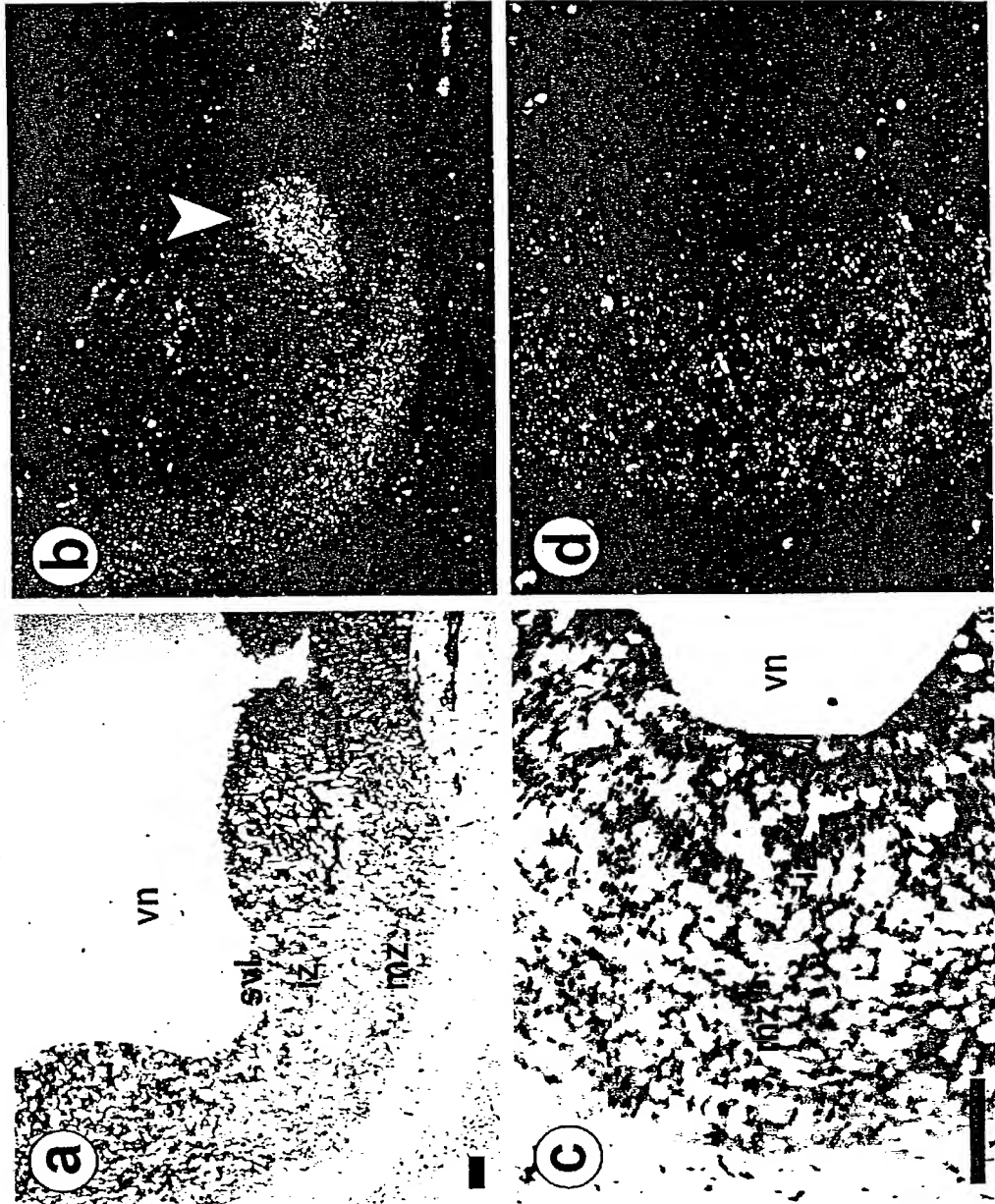
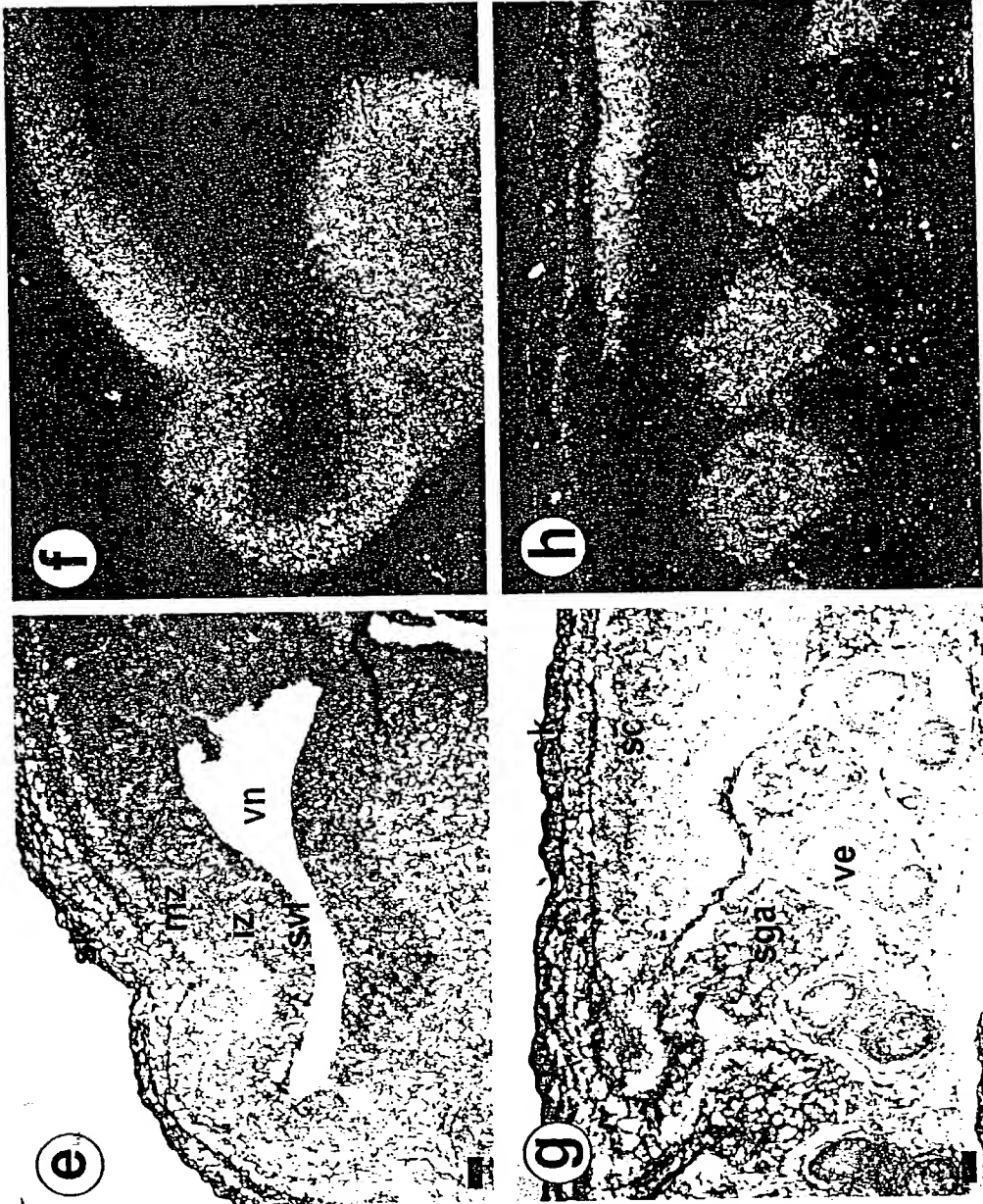


Fig. 28

122/124

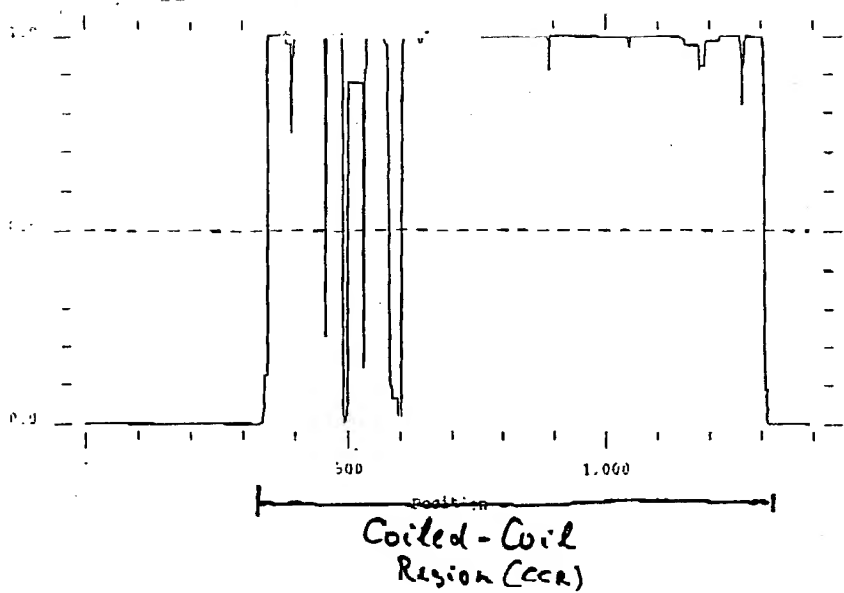


0000143113-000013112

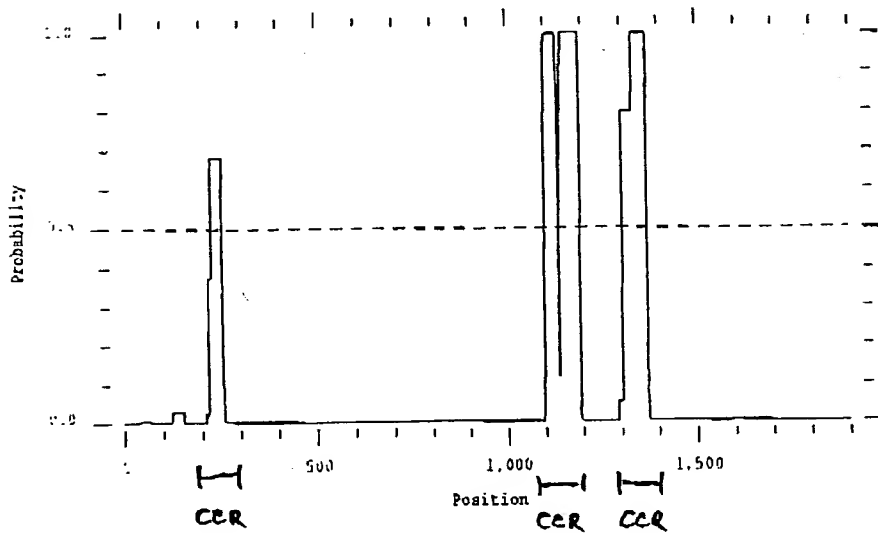
09/914549  
123/124

Fig. 29

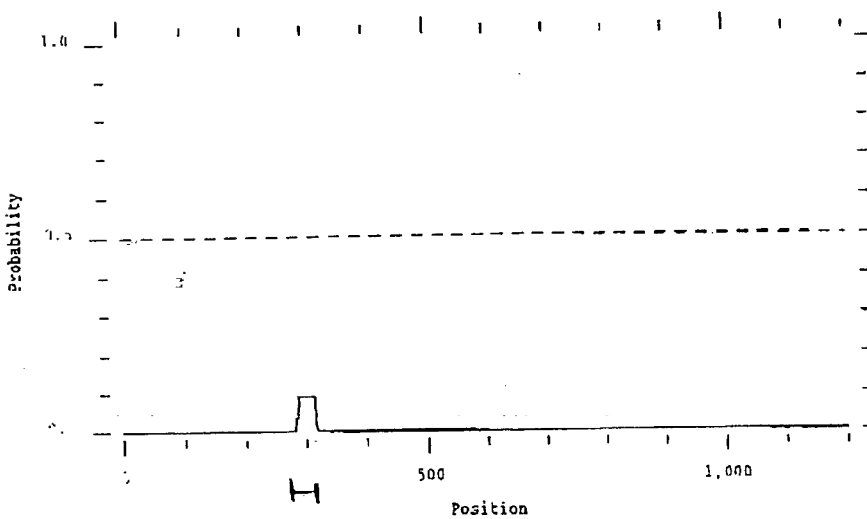
CLIP-170  
Probability



T-Protein  
Probability



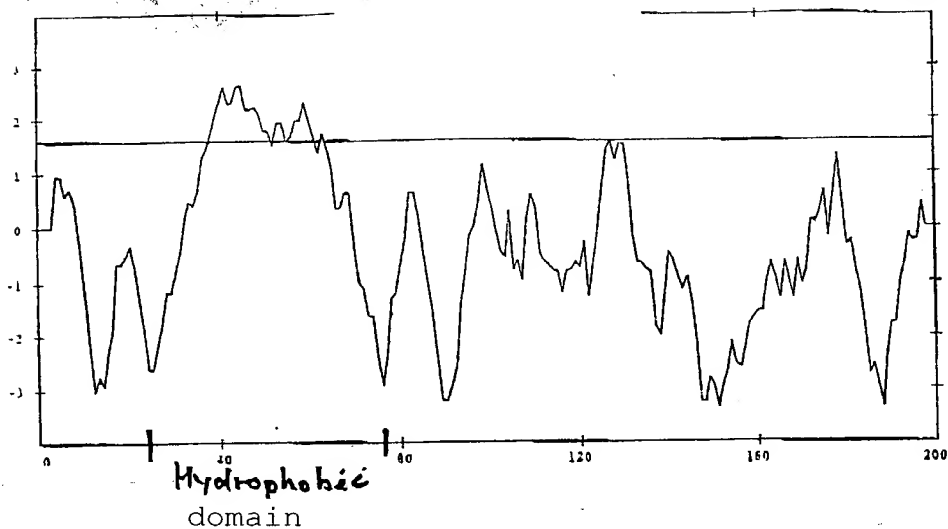
POM 121  
Probability



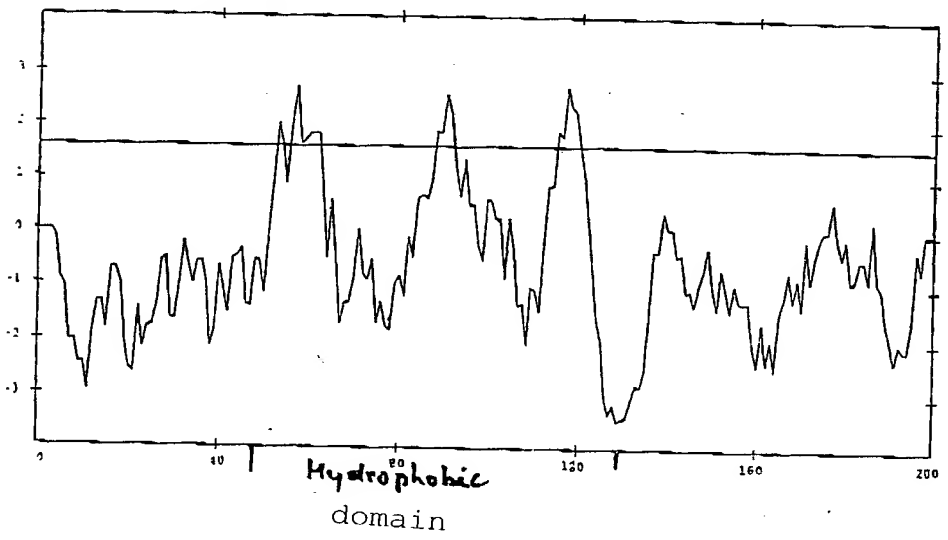
124/124

Fig. 30

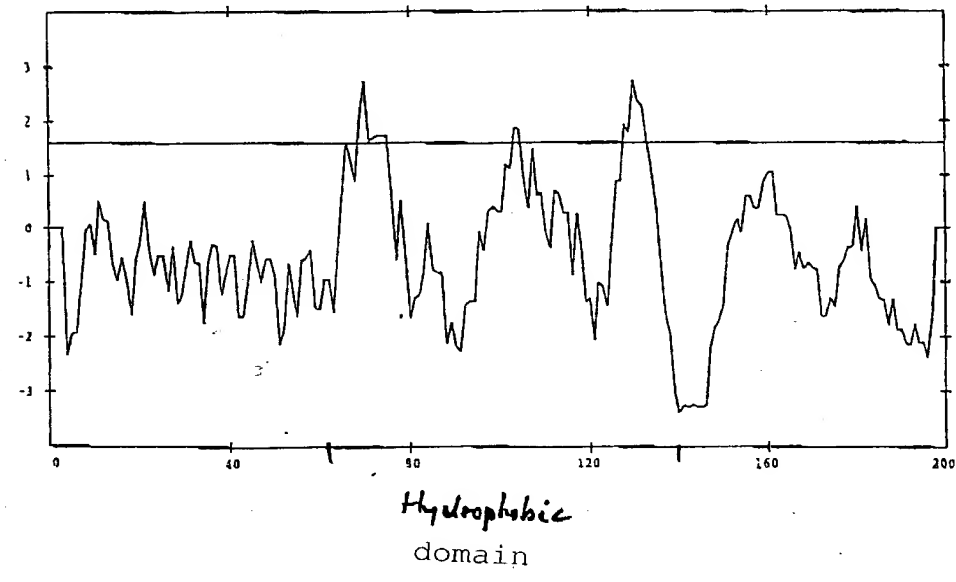
PM121



T-Protein



T3-Protein



## PATENT APPLICATION

DECLARATION AND POWER OF ATTORNEY  
FOR PATENT APPLICATION

ATTORNEY DOCKET NO. 4121-129

As a below named inventor, I hereby declare that:

My residence/post office address and citizenship are as stated below next to my name;

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter, which is claimed and for which a patent is sought on the invention entitled:

PROTEIN (TP) THAT IS INVOLVED IN THE DEVELOPMENT OF THE NERVOUS SYSTEM

the specification of which is attached hereto unless the following box is checked:

(X) was filed on August 24, 2001 as US Application Serial No. 09/914,549 or PCT International Application

Number \_\_\_\_\_ and was amended on \_\_\_\_\_ (if applicable).

I hereby state that I have reviewed and understood the contents of the above-identified specification, including the claims, as amended by any amendment(s) referred to above. I acknowledge the duty to disclose all information which is material to patentability as defined in 37 CFR 1.56.

## Foreign Application(s) and/or Claim of Foreign Priority

I hereby claim foreign priority benefits under Title 35, United States Code Section 119(a-d) or 365(b) of any foreign application(s) for patent or inventor(s) certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below any foreign application for patent or inventor(s) certificate having a filing date before that of the application on which priority is claimed:

COUNTRY	APPLICATION NUMBER	DATE FILED	PRIORITY CLAIMED UNDER 35 U.S.C. 119
Germany	199 08 423.8	26 February 1999	YES: <input checked="" type="checkbox"/> NO: <input type="checkbox"/>
PCT	PCT/DE00/00583	28 February 2000	YES: <input checked="" type="checkbox"/> NO: <input type="checkbox"/>

## Provisional Application

I hereby claim the benefit under Title 35, United States Code Section 119(e) of any United States provisional application(s) listed below:

## U.S. Priority Claim

I hereby claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claim of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

APPLICATION SERIAL NUMBER	FILING DATE	STATUS(patented/pending/abandoned)

## POWER OF ATTORNEY:

As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) listed below to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

Steven J. Hultquist, Reg. No. 28,021

Marianne Fuierer, Reg. No. 39,983

Send Correspondence to:	Direct Telephone Calls To:
Steven J. Hultquist Intellectual Property/Technology Law P.O. Box 14329 Research Triangle Park, NC 27709	Steven J. Hultquist (919) 419-9350

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full Name of Inventor: Annemarie Poustka

Citizenship: Austrian

Residence: Wendelsboos 36, D-69120 Heidelberg, Germany

Post Office Address: Same

Inventor's Signature

Date

30. Okt. 01

09914543 050502

DECLARATION AND POWER OF ATTORNEY  
FOR PATENT APPLICATION

ATTORNEY DOCKET NO. 4121-129

200  
Full Name of Inventor: Johannes Coy

Citizenship: German

Residence: In den Schwarzen Garten 1, D-63762 Grossostheim, Germany

Post Office Address: Same

Johannes Coy  
Inventor's Signature

Date

30.10.01



SEQUENCE LISTING

<110> Poustka, et al.

<120> Protein (TP) That is Involved in the Development of the Nervous System

<130> 4121-129

<140>

<141>

<150> PCT/DE00/00583  
DE 199 08 423.8

<151> 2000-02-28  
1999-02-26

<160> 39

<170> PatentIn Ver. 2.0

<210> 1

<211> 242

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<223> (3)..(242)

<400> 1

ag cga gtt act cac gct tcc cct cca tgc gaa gcc agc cag gcc aaa	47
Arg Val Thr His Ala Ser Pro Pro Ser Glu Ala Ser Gln Ala Lys	
1 5 10 15	
acc cag caa gat atg cag tcc agt ctg gca gcc aga tat gca act cag	95
Thr Gln Gln Asp Met Gln Ser Ser Leu Ala Ala Arg Tyr Ala Thr Gln	
20 25 30	
tct aat cac agt gga att gca acc agt caa aaa aag cct act agg ctt	143
Ser Asn His Ser Gly Ile Ala Thr Ser Gln Lys Lys Pro Thr Arg Leu	
35 40 45	
cca ggg ccc tct agg gtg cct gct gca gga agc agc agc aag gtc cag	191
Pro Gly Pro Ser Arg Val Pro Ala Ala Gly Ser Ser Ser Lys Val Gln	
50 55 60	
gga gcc tct aat tta aat agg aga agt cag agc ttt aac agc att gac	239
Gly Ala Ser Asn Leu Asn Arg Arg Ser Gln Ser Phe Asn Ser Ile Asp	
65 70 75	
aaa	242
Lys	

80

<210> 2  
 <211> 80  
 <212> PRT  
 <213> Homo sapiens

&lt;400&gt; 2

Arg	Val	Thr	His	Ala	Ser	Pro	Pro	Ser	Glu	Ala	Ser	Gln	Ala	Lys	Thr
1				5					10					15	
Gln	Gln	Asp	Met	Gln	Ser	Ser	Leu	Ala	Ala	Arg	Tyr	Ala	Thr	Gln	Ser
		20						25					30		
Asn	His	Ser	Gly	Ile	Ala	Thr	Ser	Gln	Lys	Lys	Pro	Thr	Arg	Leu	Pro
		35					40					45			
Gly	Pro	Ser	Arg	Val	Pro	Ala	Ala	Gly	Ser	Ser	Ser	Lys	Val	Gln	Gly
	50					55					60				
Ala	Ser	Asn	Leu	Asn	Arg	Arg	Ser	Gln	Ser	Phe	Asn	Ser	Ile	Asp	Lys
65					70					75					80

<210> 3  
 <211> 159  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <223> (1)..(159)

&lt;400&gt; 3

ggc	act	cac	gag	gtc	cag	agc	ctg	ctc	atg	aga	acg	ggt	agt	gtg	aga	48
Gly	Thr	His	Glu	Val	Gln	Ser	Leu	Leu	Met	Arg	Thr	Gly	Ser	Val	Arg	
			5						10					15		
tct	act	ctc	tca	gaa	aga	tat	acc	cca	tca	tct	cgg	cag	gcc	aac	caa	96
Ser	Thr	Leu	Ser	Glu	Arg	Tyr	Thr	Pro	Ser	Ser	Arg	Gln	Ala	Asn	Gln	
		20						25					30			
gaa	gag	ggc	aaa	gag	tgg	ttg	cgt	tct	cat	tct	act	gga	ggg	ctt	cag	144
Glu	Glu	Gly	Lys	Glu	Trp	Leu	Arg	Ser	His	Ser	Thr	Gly	Gly	Leu	Gln	
		35				40						45				
gac	act	ggc	aac	cag												159
Asp	Thr	Gly	Asn	Gln												
		50														

<210> 4  
 <211> 53  
 <212> PRT  
 <213> Homo sapiens

<400> 4

Gly Thr His Glu Val Gln Ser Leu Leu Met Arg Thr Gly Ser Val Arg  
 1 5 10 15  
 Ser Thr Leu Ser Glu Arg Tyr Thr Pro Ser Ser Arg Gln Ala Asn Gln  
 20 25 30  
 Glu Glu Gly Lys Glu Trp Leu Arg Ser His Ser Thr Gly Gly Leu Gln  
 35 40 45  
 Asp Thr Gly Asn Gln  
 50

<210> 5  
 <211> 2461  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <223> (2)..(1627)

<400> 5

g gat cag ctt cgg gag acc atg cac aac atg cag ttg gag gtg gac 46  
 Asp Gln Leu Arg Glu Thr Met His Asn Met Gln Leu Glu Val Asp  
 ctg ctg aaa gca gag aat gac cga ctg aag gta gcc cca ggc ccc tca 94  
 Leu Leu Lys Ala Glu Asn Asp Arg Leu Lys Val Ala Pro Gly Pro Ser  
 tca ggc tcc act cca ggg cag gtc cct gga tca tct gca tta tct tcc 142  
 Ser Gly Ser Thr Pro Gly Gln Val Pro Gly Ser Ser Ala Leu Ser Ser  
 cca cgc cgc tcc cta ggc ctg gca ctc acc cat tcc ttc ggc ccc agt 190  
 Pro Arg Arg Ser Leu Gly Leu Ala Leu Thr His Ser Phe Gly Pro Ser  
 ctt gca gac aca gac ctg tca ccc atg gat ggc atc agt act tgt ggt 238  
 Leu Ala Asp Thr Asp Leu Ser Pro Met Asp Gly Ile Ser Thr Cys Gly  
 cca aag gag gaa gtg acc ctc cgg gtg gtg gtg agg atg ccc ccg cag 286  
 Pro Lys Glu Glu Val Thr Leu Arg Val Val Val Arg Met Pro Pro Gln

cac atc atc aaa ggg gac ttg aag cag cag gaa ttc ttc ctg ggc tgt His Ile Ile Lys Gly Asp Leu Lys Gln Gln Glu Phe Phe Leu Gly Cys	334
agc aag gtc agt gga aaa gtt gac tgg aag atg ctg gat gaa gct gtt Ser Lys Val Ser Gly Lys Val Asp Trp Lys Met Leu Asp Glu Ala Val	382
ttc caa gtg ttc aag gac tat att tct aaa atg gac cca gcc tct acc Phe Gln Val Phe Lys Asp Tyr Ile Ser Lys Met Asp Pro Ala Ser Thr	430
ctg gga cta agc act gag tcc atc cat ggc tac agc atc agc cac gtg Leu Gly Leu Ser Thr Glu Ser Ile His Gly Tyr Ser Ile Ser His Val	478
aaa cga gtg ttg gat gca gag ccc ccc gag atg cct cct tgc cgt cga Lys Arg Val Leu Asp Ala Glu Pro Pro Glu Met Pro Pro Cys Arg Arg	526
ggc gtc aat aac ata tca gtc tcc ctc aaa ggt ctg aag gag aaa tgc Gly Val Asn Asn Ile Ser Val Ser Leu Lys Gly Leu Lys Glu Lys Cys	574
gtc gac agc ctg gtg ttc gag acg ctg atc ccc aag ccg atg atg cag Val Asp Ser Leu Val Phe Glu Thr Leu Ile Pro Lys Pro Met Met Gln	622
cac tac ata agc ctc ctg ctg aag cac cgg cgc ctc gtc ctc tgc ggc His Tyr Ile Ser Leu Leu Leu Lys His Arg Arg Leu Val Leu Ser Gly	670
ccc agc ggc acg ggc aag acc tac ctg acc aat cgc ttg gcc gag tac Pro Ser Gly Thr Gly Lys Thr Tyr Leu Thr Asn Arg Leu Ala Glu Tyr	718
ctg gtg gag cgc tct ggc cgt gag gtc aca gag ggc atc gtc agc acc Leu Val Glu Arg Ser Gly Arg Glu Val Thr Glu Gly Ile Val Ser Thr	766
ttc aac atg cac cag cag tct tgc aag gat ctg caa ctg tat ctt tcc Phe Asn Met His Gln Gln Ser Cys Lys Asp Leu Gln Leu Tyr Leu Ser	814
aac cta gcc aac cag ata gac cgg gaa aca gga att ggg gat gtg ccc Asn Leu Ala Asn Gln Ile Asp Arg Glu Thr Gly Ile Gly Asp Val Pro	862
ctg gtg att cta ttg gat gac ctg agt gaa gca ggc tcc atc agt gag Leu Val Ile Leu Leu Asp Asp Leu Ser Glu Ala Gly Ser Ile Ser Glu	910
ttg gtc aat ggg gcc ctc acc tgc aag tat cat aaa tgt ccc tat att Leu Val Asn Gly Ala Leu Thr Cys Lys Tyr His Lys Cys Pro Tyr Ile	958

5



```

taagggttcg gcaatcactg tcacccccgg acagcagaac gctggcatca gctatcttag      1687
ctcctcctct cccctctcct ctttcagagc actggctctc cagccccagg aggagaacag      1747
gagggaggag gagatgaaag aggagggaca ggttcttggg gctgtacctt tgagaacttc      1807
ctaggaagga atggtggggg ggcgtttggg aacttgtgcc ccctaaacac atttactggc      1867
ctcctctaata gactttgggg aaaagatgat tctgggtctt tcccttgact tcttgtttca      1927
attacaaact cctgggcttt ctggggaggg gttcagaaaa catcaaaaca ctgcagcagt      1987
tcctaaatga ttctcacaag caaccctgag agagacagtc ttgtgagggg gatctggggg      2047
aggcaggaag ctcctcagat tttctcacag acccttccca attccatcac cactgccaac      2107
aactcctccc ccagagatct ggctggagcc cagaaaaaga agcatgtggg ttaaaaaatg      2167
tttaaatcaa tctgtaaaag gtaaaaatga aaaacaaaaa caagcaaaca aacaaaaaac      2227
aatggaaaag atgaagctgg agagagagga accagttgcc aaggtagaga gctgcccgt      2287
cctgccctct ggatgacata ggggacatca acaagacggc tgccaacctg agaagtcacc      2347
aaaccacaaa aataacctta cagccttcag ggaaagacta ccagctctgt ctttctaccc      2407
tctaatttaa caatgcataa gagtcaataa accctacttt tttaaaaaaa aaaa          2461

```

<210> 6  
 <211> 542  
 <212> PRT  
 <213> Homo sapiens

<400> 6

```

Asp Gln Leu Arg Glu Thr Met His Asn Met Gln Leu Glu Val Asp Leu
 1              5              10              15
Leu Lys Ala Glu Asn Asp Arg Leu Lys Val Ala Pro Gly Pro Ser Ser
          20              25              30
Gly Ser Thr Pro Gly Gln Val Pro Gly Ser Ser Ala Leu Ser Ser Pro
          35              40              45
Arg Arg Ser Leu Gly Leu Ala Leu Thr His Ser Phe Gly Pro Ser Leu
          50              55              60
Ala Asp Thr Asp Leu Ser Pro Met Asp Gly Ile Ser Thr Cys Gly Pro
 65              70              75              80
Lys Glu Glu Val Thr Leu Arg Val Val Val Arg Met Pro Pro Gln His
          85              90              95
Ile Ile Lys Gly Asp Leu Lys Gln Gln Glu Phe Phe Leu Gly Cys Ser

```

100					105					110						
Lys	Val	Ser	Gly	Lys	Val	Asp	Trp	Lys	Met	Leu	Asp	Glu	Ala	Val	Phe	
115					120					125						
Gln	Val	Phe	Lys	Asp	Tyr	Ile	Ser	Lys	Met	Asp	Pro	Ala	Ser	Thr	Leu	
130					135					140						
Gly	Leu	Ser	Thr	Glu	Ser	Ile	His	Gly	Tyr	Ser	Ile	Ser	His	Val	Lys	
145					150					155					160	
Arg	Val	Leu	Asp	Ala	Glu	Pro	Pro	Glu	Met	Pro	Pro	Cys	Arg	Arg	Gly	
165					170					175						
Val	Asn	Asn	Ile	Ser	Val	Ser	Leu	Lys	Gly	Leu	Lys	Glu	Lys	Cys	Val	
180					185					190						
Asp	Ser	Leu	Val	Phe	Glu	Thr	Leu	Ile	Pro	Lys	Pro	Met	Met	Gln	His	
195					200					205						
Tyr	Ile	Ser	Leu	Leu	Leu	Lys	His	Arg	Arg	Leu	Val	Leu	Ser	Gly	Pro	
210					215					220						
Ser	Gly	Thr	Gly	Lys	Thr	Tyr	Leu	Thr	Asn	Arg	Leu	Ala	Glu	Tyr	Leu	
225					230					235					240	
Val	Glu	Arg	Ser	Gly	Arg	Glu	Val	Thr	Glu	Gly	Ile	Val	Ser	Thr	Phe	
245					250					255						
Asn	Met	His	Gln	Gln	Ser	Cys	Lys	Asp	Leu	Gln	Leu	Tyr	Leu	Ser	Asn	
260					265					270						
Leu	Ala	Asn	Gln	Ile	Asp	Arg	Glu	Thr	Gly	Ile	Gly	Asp	Val	Pro	Leu	
275					280					285						
Val	Ile	Leu	Leu	Asp	Asp	Leu	Ser	Glu	Ala	Gly	Ser	Ile	Ser	Glu	Leu	
290					295					300						
Val	Asn	Gly	Ala	Leu	Thr	Cys	Lys	Tyr	His	Lys	Cys	Pro	Tyr	Ile	Ile	
305					310					315					320	
Gly	Thr	Thr	Asn	Gln	Pro	Val	Lys	Met	Thr	Pro	Asn	His	Gly	Leu	His	
325					330					335						
Leu	Ser	Phe	Arg	Met	Leu	Thr	Phe	Ser	Asn	Asn	Val	Glu	Pro	Ala	Asn	
340					345					350						
Gly	Phe	Leu	Val	Arg	Tyr	Leu	Arg	Arg	Lys	Leu	Val	Glu	Ser	Asp	Ser	
355					360					365						
Asp	Ile	Asn	Ala	Asn	Lys	Glu	Glu	Leu	Leu	Arg	Val	Leu	Asp	Trp	Val	
370					375					380						
Pro	Lys	Leu	Trp	Tyr	His	Leu	His	Thr	Phe	Leu	Glu	Lys	His	Ser	Thr	
385					390					395					400	
Ser	Asp	Phe	Leu	Ile	Gly	Pro	Cys	Phe	Phe	Leu	Ser	Cys	Pro	Ile	Gly	

405	410	415
Ile Glu Asp Phe Arg Thr Trp Phe Ile Asp Leu Trp Asn Asn Ser Ile 420 425 430		
Ile Pro Tyr Leu Gln Glu Gly Ala Lys Asp Gly Ile Lys Val His Gly 435 440 445		
Gln Lys Ala Ala Trp Glu Asp Pro Val Glu Trp Val Arg Asp Thr Leu 450 455 460		
Pro Trp Pro Ser Ala Gln Gln Asp Gln Ser Lys Leu Tyr His Leu Pro 465 470 475 480		
Pro Pro Thr Val Gly Pro His Ser Ile Ala Ser Pro Pro Glu Asp Arg 485 490 495		
Thr Val Lys Asp Ser Thr Pro Ser Ser Leu Asp Ser Asp Pro Leu Met 500 505 510		
Ala Met Leu Leu Lys Leu Gln Glu Ala Ala Asn Tyr Ile Glu Ser Pro 515 520 525		
Asp Arg Glu Thr Ile Leu Asp Pro Asn Leu Gln Ala Thr Leu 530 535 540		

<210> 7  
 <211> 2568  
 <212> DNA  
 <213> mouse

<220>  
 <221> CDS  
 <223> (1)..(1695)

<400> 7

gaa cta tgg gaa aaa gag atg aag ctc acg gat atc cgg ttg gag gcc Glu Leu Trp Glu Lys Glu Met Lys Leu Thr Asp Ile Arg Leu Glu Ala	48
ctc aac tct gcc cac cag ctg gac cag ctt cgg gag acc atg cac aat Leu Asn Ser Ala His Gln Leu Asp Gln Leu Arg Glu Thr Met His Asn	96
atg cag ttg gag gtg gac ctg ctg aaa gca gag aat gac cgg ctg aag Met Gln Leu Glu Val Asp Leu Leu Lys Ala Glu Asn Asp Arg Leu Lys	144
gtt gcc ccc ggc ccc tcc tca ggc tgc act cca ggg cag gtc cct ggg Val Ala Pro Gly Pro Ser Ser Gly Cys Thr Pro Gly Gln Val Pro Gly	192



tca tcg gct ctg tgg tcc cct cga cgt tcc ctg ggc ctt gca ctc agc Ser Ser Ala Leu Ser Ser Pro Arg Arg Ser Leu Gly Leu Ala Leu Ser	240
cat cct ttc agt cct agt ctc aca gac aca gac ctc tca ccc atg gat His Pro Phe Ser Pro Ser Leu Thr Asp Thr Asp Leu Ser Pro Met Asp	288
ggc atc agc acc tgt ggt tca aag gaa gag gtg acc ctg cgg gtg gtg Gly Ile Ser Thr Cys Gly Ser Lys Glu Glu Val Thr Leu Arg Val Val	336
gtc cgg atg ccg ccc cag cac atc atc aaa ggg gac tta aag cag cag Val Arg Met Pro Pro Gln His Ile Ile Lys Gly Asp Leu Lys Gln Gln	384
gag ttc ttc ctg ggt tgc agc aag gtc agt ggc aaa gtt gac tgg aag Glu Phe Phe Leu Gly Cys Ser Lys Val Ser Gly Lys Val Asp Trp Lys	432
atg ctg gat gaa gcc gtt ttc caa gtg ttc aag gac tac att tct aaa Met Leu Asp Glu Ala Val Phe Gln Val Phe Lys Asp Tyr Ile Ser Lys	480
atg gac cca gcc tca acc ctg gga ctg agc act gag tcc ata cat ggc Met Asp Pro Ala Ser Thr Leu Gly Leu Ser Thr Glu Ser Ile His Gly	528
tat agc ctc agc cac gtg aaa cga gtg ctg gat gct gag ccc cca gag Tyr Ser Leu Ser His Val Lys Arg Val Leu Asp Ala Glu Pro Pro Glu	576
atg cct cct tgc cgc cga ggt gtc aat aac ata tca gtc gct ctc aaa Met Pro Pro Cys Arg Arg Gly Val Asn Asn Ile Ser Val Ala Leu Lys	624
ggc ctg aaa gag aag tgt gtc gac agc ctg gtg ttc gag acg ctt atc Gly Leu Lys Glu Lys Cys Val Asp Ser Leu Val Phe Glu Thr Leu Ile	672
ccc aag ccc atg atg cag cac tac atc agc ctc ctg ctc aag cac cgg Pro Lys Pro Met Met Gln His Tyr Ile Ser Leu Leu Leu Lys His Arg	720
cgc ctg gtg ctc tcc ggc ccc agt ggc acc ggc aag acc tac ttg acc Arg Leu Val Leu Ser Gly Pro Ser Gly Thr Gly Lys Thr Tyr Leu Thr	768
aat cgg cta gcc gag tac ctg gtg gag cgc tcc ggc cgc gag gtc acg Asn Arg Leu Ala Glu Tyr Leu Val Glu Arg Ser Gly Arg Glu Val Thr	816
gat ggc atc gtc agc act ttc aac atg cac cag cag tct tgc aag gat Asp Gly Ile Val Ser Thr Phe Asn Met His Gln Gln Ser Cys Lys Asp	864

ctg caa ctg tac ctc tcc aac cta gcc aac cag ata gac cgg gaa aca Leu Gln Leu Tyr Leu Ser Asn Leu Ala Asn Gln Ile Asp Arg Glu Thr	912
ggg ata ggg gat gtg ccc ttg gtg atc ctc ctg gat gat ctg agt gaa Gly Ile Gly Asp Val Pro Leu Val Ile Leu Leu Asp Asp Leu Ser Glu	960
gca ggc tcc atc agt gag ctg gtc aat ggg gcc ctc acc tgc aag tat Ala Gly Ser Ile Ser Glu Leu Val Asn Gly Ala Leu Thr Cys Lys Tyr	1008
cac aaa tgt ccc tac att ata ggt acc acc aat cag cct gta aaa atg His Lys Cys Pro Tyr Ile Ile Gly Thr Thr Asn Gln Pro Val Lys Met	1056
aca ccc aac cat ggc ttg cac ttg agc ttc agg atg ctg acc ttc tcg Thr Pro Asn His Gly Leu His Leu Ser Phe Arg Met Leu Thr Phe Ser	1104
aac aat gtg gaa cca gcc aat ggc ttt ctg gtc cgt tac ctg cgg agg Asn Asn Val Glu Pro Ala Asn Gly Phe Leu Val Arg Tyr Leu Arg Arg	1152
aag ttg gta gag tca gac agt gac gtc aat gct aac aag gaa gag ctg Lys Leu Val Glu Ser Asp Ser Asp Val Asn Ala Asn Lys Glu Glu Leu	1200
ctt cgg gtg ctg gac tgg gtg ccc aag ctg tgg tat cac ctc cac acc Leu Arg Val Leu Asp Trp Val Pro Lys Leu Trp Tyr His Leu His Thr	1248
ttc ctg gag aag cac agc acc tcg gac ttc ctc att ggc cct tgc ttc Phe Leu Glu Lys His Ser Thr Ser Asp Phe Leu Ile Gly Pro Cys Phe	1296
ttc ctg tcc tgt ccc att ggc atc gag gac ttc cgg acc tgg ttc att Phe Leu Ser Cys Pro Ile Gly Ile Glu Asp Phe Arg Thr Trp Phe Ile	1344
gac ctg tgg aac aat tcc atc atc ccc tat cta cag gaa gga gcc aag Asp Leu Trp Asn Asn Ser Ile Ile Pro Tyr Leu Gln Glu Gly Ala Lys	1392
gat ggg atc aag gtt cat gga cag aaa gct gct tgg gaa gac ccg gtg Asp Gly Ile Lys Val His Gly Gln Lys Ala Ala Trp Glu Asp Pro Val	1440
gaa tgg gtc cga gac act ctt ccc tgg ccg tcg gcc caa caa gac caa Glu Trp Val Arg Asp Thr Leu Pro Trp Pro Ser Ala Gln Gln Asp Gln	1488
tca aag ctc tac cac ctg ccc ccg cct tct gtg ggc ccc cac agc act Ser Lys Leu Tyr His Leu Pro Pro Pro Ser Val Gly Pro His Ser Thr	1536

gcc tca ccc ccg gag gac agg aca gtc aaa gac agc act cca aac tcc 1584  
Ala Ser Pro Pro Glu Asp Arg Thr Val Lys Asp Ser Thr Pro Asn Ser

ctc gac tca gat ccc ctg atg gcc atg cta ctg aaa ctc caa gaa gct 1632  
Leu Asp Ser Asp Pro Leu Met Ala Met Leu Leu Lys Leu Gln Glu Ala

gcc aac tac att gag tca cca gat cga gag act atc ctg gac ccc aac 1680  
Ala Asn Tyr Ile Glu Ser Pro Asp Arg Glu Thr Ile Leu Asp Pro Asn

ctc cag gcg aca ctc tgagggcccg gcagtcactg tcaccctgga gggcagaagg 1735  
Leu Gln Ala Thr Leu

ctggcttcag catcattagc tctctctgc cctcttctt catagctctg gctcaccagc 1795

ctcgccaaga gaacaggagg gaagaagagg gcaggaggag ggatgggttc tcggtgctga 1855

acctttgaga acttcctact aggaattgga ggggggtggag tttgagaact ccgtgcccct 1915

taactacatt tgctggcctc ctcttacgac ttaggagaaa agatgattct ggtcttttct 1975

tcaagttttg tttcacctac aaactcttgg gctttctggg gagggattcg gaagatataa 2035

acagacaaac aaaaacaaac aaaccaacta cagcagttcc aagctcgttc tcacaaacac 2095

ctctgagaca gtcacatgtg ggcaaatacta agggaggcag gaagctctac agactttctt 2155

gcaaaccctt cccagttctg tcgacactgc caacaacctc cccgccagag acctggccag 2215

agccaagaaa agagaagcat gtggtttaac agaaaaacaa aacaaaacaa aacaaaaaat 2275

atatgtgtaa atcaacctgt agaaggtaaa aacggcaatg gaaaagatga agctggaagg 2335

aggggcccag ttgccaagat ggaacgagag ctgccagatc ttgccttctg gatgacaaga 2395

ggggacattg caagatggct gccagtctaa aacgtcacca gaccacaaga gtaacatcac 2455

agccttcgaa gaaaggccac aagctgtctt tctgccctct aactgaacat gcatgaaaag 2515

tcaataaacc ctacttttta atttttaaaa aaaaaaaaaa aaaaaaaaaa aaa 2568

```
<210> 8
<211> 565
<212> PRT
<213> mouse
```

<400> 8

Glu Leu Trp Glu Lys Glu Met Lys Leu Thr Asp Ile Arg Leu Glu Ala  
1 5 10 15

Leu Asn Ser Ala His Gln Leu Asp Gln Leu Arg Glu Thr Met His Asn  
20 25 30

Met Gln Leu Glu Val Asp Leu Leu Lys Ala Glu Asn Asp Arg Leu Lys  
 35 40 45  
 Val Ala Pro Gly Pro Ser Ser Gly Cys Thr Pro Gly Gln Val Pro Gly  
 50 55 60  
 Ser Ser Ala Leu Ser Ser Pro Arg Arg Ser Leu Gly Leu Ala Leu Ser  
 65 70 75 80  
 His Pro Phe Ser Pro Ser Leu Thr Asp Thr Asp Leu Ser Pro Met Asp  
 85 90 95  
 Gly Ile Ser Thr Cys Gly Ser Lys Glu Glu Val Thr Leu Arg Val Val  
 100 105 110  
 Val Arg Met Pro Pro Gln His Ile Ile Lys Gly Asp Leu Lys Gln Gln  
 115 120 125  
 Glu Phe Phe Leu Gly Cys Ser Lys Val Ser Gly Lys Val Asp Trp Lys  
 130 135 140  
 Met Leu Asp Glu Ala Val Phe Gln Val Phe Lys Asp Tyr Ile Ser Lys  
 145 150 155 160  
 Met Asp Pro Ala Ser Thr Leu Gly Leu Ser Thr Glu Ser Ile His Gly  
 165 170 175  
 Tyr Ser Leu Ser His Val Lys Arg Val Leu Asp Ala Glu Pro Pro Glu  
 180 185 190  
 Met Pro Pro Cys Arg Arg Gly Val Asn Asn Ile Ser Val Ala Leu Lys  
 195 200 205  
 Gly Leu Lys Glu Lys Cys Val Asp Ser Leu Val Phe Glu Thr Leu Ile  
 210 215 220  
 Pro Lys Pro Met Met Gln His Tyr Ile Ser Leu Leu Leu Lys His Arg  
 225 230 235 240  
 Arg Leu Val Leu Ser Gly Pro Ser Gly Thr Gly Lys Thr Tyr Leu Thr  
 245 250 255  
 Asn Arg Leu Ala Glu Tyr Leu Val Glu Arg Ser Gly Arg Glu Val Thr  
 260 265 270  
 Asp Gly Ile Val Ser Thr Phe Asn Met His Gln Gln Ser Cys Lys Asp  
 275 280 285  
 Leu Gln Leu Tyr Leu Ser Asn Leu Ala Asn Gln Ile Asp Arg Glu Thr  
 290 295 300  
 Gly Ile Gly Asp Val Pro Leu Val Ile Leu Leu Asp Asp Leu Ser Glu  
 305 310 315 320  
 Ala Gly Ser Ile Ser Glu Leu Val Asn Gly Ala Leu Thr Cys Lys Tyr

[illegible]

```
<210> 9
<211> 1025
<212> DNA
<213> mouse
```

```
<220>  
<221> CDS  
<223> (1) .. (129)
```

&lt;400&gt; 9

```

cca ata gaa ctc cgg atc aag agg cag aat tcc tca gat agc atc tcc      48
Pro Ile Glu Leu Arg Ile Lys Arg Gln Asn Ser Ser Asp Ser Ile Ser

agc ctc aac agc atc acc agc cat tcc agc atc ggc agc agc aaa gat      96
Ser Leu Asn Ser Ile Thr Ser His Ser Ser Ile Gly Ser Ser Lys Asp

gct gat gcc aag aag aaa aag aag aag agt tgg gtatgtaaag gcttggggat    149
Ala Asp Ala Lys Lys Lys Lys Lys Lys Ser Trp

cggcctgtgc taggagtcac tcaccctgtt gcagggaact gacccctttc aggatcaaca    209

aagaggggtcc cttctaacag gatgccagtg ttgtgacatc tgctggggac aaaaattcac    269

taagtcccca ttcctctatc cattgtctat tctccttacc accgccctgc acatataccc    329

cagcccccca ccgtccctgc atcctttata catgtctgct atcctggggc tctacctact    389

gatgaggcca aatgtatttg gccgtagaag gagctgagaa aattattcat ggggtgggaga    449

gtggggcatg tggagagaat ttgtaagcca agcagggtac tctagacgct cctggggctg     509

ttgcttttagt ttgggtgagg aggctgtgga acgtcccatc cgctccaaag cctgcttttg    569

tctgggccag aggtgggttt gttctgtgtg gtatcccccc tgtaactcta aactggcttt     629

gggtgagctt tctacaatct gtacgcaggt gtagggcact gcctgactga ctgaaaggga     689

gagtgacca gagtgagcgt cttgtccctg tccctgctga ggagggtgg ctacagactt     749

tggcctagtg cagacaggag ccagctgtgt ggagaagcag ctgtgtgaaa tgcattgagta    809

gtgtcgtgc tgctgctgct gctgctttct tttcattggt tttttttttt tttctttcct    869

tttatttcct tcaaaatgct gacctcaaat ccctattttt tttccagggt tatgaggtaa    929

gaaactcggg tcctctctc gtgctttttc tttccctttg cacaccttcg tgtttccaga    989

gcaagcacct ctcttaaaaa aaaaaaaaaa aaaaaa                               1025

```

<210> 10  
 <211> 43  
 <212> PRT  
 <213> mouse

&lt;400&gt; 10

```

Pro Ile Glu Leu Arg Ile Lys Arg Gln Asn Ser Ser Asp Ser Ile Ser
  1             5             10             15
Ser Leu Asn Ser Ile Thr Ser His Ser Ser Ile Gly Ser Ser Lys Asp

```

20

25

30

Ala Asp Ala Lys Lys Lys Lys Lys Lys Ser Trp  
 35 40

<210> 11  
 <211> 8690  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <223> (503)..(6187)

<400> 11

```

ctctcagcgg cccaagcagt ttctttctgg gtgacaagaa tgtgcctcgg ttggtttttc      60
ttttttttct ccatctcctt aagacgattt ccatagtaac ctgatcaagt ggctcaaaat      120
cgcaaacctg aggatttccg cggcccgccg gcaagacctc ggccagggtt acaggaaatc      180
tgtcattttt tattaaaaatg gaaaactgtg aagaaagaaa aagatagcag ttgaagtcaa      240
aattctcgga tgactatttt gcttttgagg agtcagcatt taaaaacgat atgctgattt      300
ggaaggtcct gggagtaaac tgcaaaacttt attttttcca ttcaatcaat ggatttttta      360
atcattcctt ggagtcgatg aagttcggaa acggtgtgtg atgggggaacg tggcggggcca      420
gtgtgttcct agaaattgca tcttggatta gtttgctgct tttttgaaga gattccattt      480
tgaagggcaa gaacctaatg tg atg gat tta tct tca gaa atg aac aga cat      532
                        Met Asp Leu Ser Ser Glu Met Asn Arg His

ggg aag aat cca gtg agt cac aag cta gaa gat cag aag aag att tac      580
Gly Lys Asn Pro Val Ser His Lys Leu Glu Asp Gln Lys Lys Ile Tyr

act gac tgg gcc aac cac tac cta gca aaa tca ggc cac aag cgg ctg      628
Thr Asp Trp Ala Asn His Tyr Leu Ala Lys Ser Gly His Lys Arg Leu

atc aag gac ttg caa caa gac att gca gat gga gta ctc cta gca gaa      676
Ile Lys Asp Leu Gln Gln Asp Ile Ala Asp Gly Val Leu Leu Ala Glu

atc atc cag att att gca aat gaa aaa gtt gaa gat atc aat gga tgt      724
Ile Ile Gln Ile Ile Ala Asn Glu Lys Val Glu Asp Ile Asn Gly Cys

cct aga agt cag tct cag atg att gaa aat gtt gat gtc tgc ctt agt      772
Pro Arg Ser Gln Ser Gln Met Ile Glu Asn Val Asp Val Cys Leu Ser

```

ttt cta gca gcc aga ggg gta aat gtt caa ggt cta tct gct gaa gaa Phe Leu Ala Ala Arg Gly Val Asn Val Gln Gly Leu Ser Ala Glu Glu	820
ata aga aat gga aac tta aaa gcc att cta ggg ctg ttt ttc agt tta Ile Arg Asn Gly Asn Leu Lys Ala Ile Leu Gly Leu Phe Phe Ser Leu	868
tct cgc tac aag cag caa caa cac cat caa caa cag tac tat cag tcc Ser Arg Tyr Lys Gln Gln Gln His His Gln Gln Gln Tyr Tyr Gln Ser	916
ttg gtg gaa ctt cag cag cga gtt act cac gct tcc cct cca tcg gaa Leu Val Glu Leu Gln Gln Arg Val Thr His Ala Ser Pro Pro Ser Glu	964
gcc agc cag gcc aaa acc cag caa gat atg cag tcc agg ctt cca ggg Ala Ser Gln Ala Lys Thr Gln Gln Asp Met Gln Ser Arg Leu Pro Gly	1012
ccc tct agg gtg cct gct gca gga agc agc agc aag gtc cag gga gcc Pro Ser Arg Val Pro Ala Ala Gly Ser Ser Ser Lys Val Gln Gly Ala	1060
tct aat tta aat agg aga agt cag agc ttt aac agc att gac aaa aac Ser Asn Leu Asn Arg Arg Ser Gln Ser Phe Asn Ser Ile Asp Lys Asn	1108
aag cct cca aat tat gca aat gga aac gaa aaa ggt gaa gac cct gaa Lys Pro Pro Asn Tyr Ala Asn Gly Asn Glu Lys Gly Glu Asp Pro Glu	1156
aca aga aga atg aga aca gtt aaa aac ata gca gac ttg agg cag aat Thr Arg Arg Met Arg Thr Val Lys Asn Ile Ala Asp Leu Arg Gln Asn	1204
tta gaa gag act atg tcc agt ctt cgt ggg act cag ata agc cac agc Leu Glu Glu Thr Met Ser Ser Leu Arg Gly Thr Gln Ile Ser His Ser	1252
acc ctg gag aca aca ttt gac agc act gtg aca aca gaa gtt aat gga Thr Leu Glu Thr Thr Phe Asp Ser Thr Val Thr Thr Glu Val Asn Gly	1300
agg acc ata ccc aac ttg aca agt cga ccc acc ccc atg acc tgg agg Arg Thr Ile Pro Asn Leu Thr Ser Arg Pro Thr Pro Met Thr Trp Arg	1348
ttg ggc cag gca tgt ccg cga ctt cag gcg gga gat gct ccc tcc ctg Leu Gly Gln Ala Cys Pro Arg Leu Gln Ala Gly Asp Ala Pro Ser Leu	1396
ggg gct ggc tat cct cgc agt ggt acc agt cga ttc atc cac aca gac Gly Ala Gly Tyr Pro Arg Ser Gly Thr Ser Arg Phe Ile His Thr Asp	1444



ccc	tcg	agg	ttc	atg	tat	acc	acg	cct	ctc	cgt	cga	gct	gct	gtc	tct	1492
Pro	Ser	Arg	Phe	Met	Tyr	Thr	Thr	Pro	Leu	Arg	Arg	Ala	Ala	Val	Ser	
agg	ctg	gga	aac	atg	tca	cag	att	gac	atg	agt	gag	aaa	gca	agc	agt	1540
Arg	Leu	Gly	Asn	Met	Ser	Gln	Ile	Asp	Met	Ser	Glu	Lys	Ala	Ser	Ser	
gac	ctg	gac	atg	tct	tct	gag	gtc	gat	gtg	ggt	gga	tat	atg	agt	gat	1588
Asp	Leu	Asp	Met	Ser	Ser	Glu	Val	Asp	Val	Gly	Gly	Tyr	Met	Ser	Asp	
ggt	gat	atc	ctt	ggg	aaa	agt	ctc	agg	act	gat	gac	atc	aac	agt	ggg	1636
Gly	Asp	Ile	Leu	Gly	Lys	Ser	Leu	Arg	Thr	Asp	Asp	Ile	Asn	Ser	Gly	
tac	atg	aca	gat	gga	gga	ctt	aac	cta	tat	act	aga	agt	ctg	aac	cga	1684
Tyr	Met	Thr	Asp	Gly	Gly	Leu	Asn	Leu	Tyr	Thr	Arg	Ser	Leu	Asn	Arg	
ata	cca	gac	aca	gca	act	tcc	cgg	gac	atc	atc	cag	aga	ggg	gtt	cac	1732
Ile	Pro	Asp	Thr	Ala	Thr	Ser	Arg	Asp	Ile	Ile	Gln	Arg	Gly	Val	His	
gat	gtg	aca	gtg	gat	gca	gac	agc	tgg	gat	gac	agc	agt	tca	gtg	agc	1780
Asp	Val	Thr	Val	Asp	Ala	Asp	Ser	Trp	Asp	Asp	Ser	Ser	Ser	Val	Ser	
agt	ggt	ctc	agt	gac	acc	ctt	gat	aac	atc	agc	act	gat	gac	ctg	aac	1828
Ser	Gly	Leu	Ser	Asp	Thr	Leu	Asp	Asn	Ile	Ser	Thr	Asp	Asp	Leu	Asn	
acc	aca	tcc	tct	gtc	agc	tct	tac	tcc	aac	atc	acc	gtc	ccc	tct	agg	1876
Thr	Thr	Ser	Ser	Val	Ser	Ser	Tyr	Ser	Asn	Ile	Thr	Val	Pro	Ser	Arg	
aag	aat	act	cag	ctg	agg	aca	gat	tca	gag	aaa	cgc	tcc	acc	aca	gac	1924
Lys	Asn	Thr	Gln	Leu	Arg	Thr	Asp	Ser	Glu	Lys	Arg	Ser	Thr	Thr	Asp	
gag	acc	tgg	gat	agt	cct	gag	gaa	ctg	aaa	aaa	cca	gaa	gaa	gat	ttt	1972
Glu	Thr	Trp	Asp	Ser	Pro	Glu	Glu	Leu	Lys	Lys	Pro	Glu	Glu	Asp	Phe	
gac	agc	cat	ggg	gat	gct	ggt	ggc	aag	tgg	aag	act	gtg	tcc	tct	gga	2020
Asp	Ser	His	Gly	Asp	Ala	Gly	Gly	Lys	Trp	Lys	Thr	Val	Ser	Ser	Gly	
ctt	cct	gaa	gac	ccc	gag	aag	gca	ggg	cag	aaa	gct	tcc	ctg	tct	gtt	2068
Leu	Pro	Glu	Asp	Pro	Glu	Lys	Ala	Gly	Gln	Lys	Ala	Ser	Leu	Ser	Val	
tca	cag	acaggt	tcc	tgg	aga	aga	ggc	atg	tct	gcc	caa	gga	ggg	gcg		2116
Ser	Gln	Thr	Gly	Ser	Trp	Arg	Arg	Gly	Met	Ser	Ala	Gln	Gly	Gly	Ala	

cca tct agg cag aaa gct gga aca agt gca ctc aaa aca ccc ggg aaa Pro Ser Arg Gln Lys Ala Gly Thr Ser Ala Leu Lys Thr Pro Gly Lys	2164
acc gat gat gcc aaa gct tct gag aaa gga aaa gct ccc cta aaa gga Thr Asp Asp Ala Lys Ala Ser Glu Lys Gly Lys Ala Pro Leu Lys Gly	2212
tca tct cta caa aga tct cct tca gat gca gga aaa agc agt gga gat Ser Ser Leu Gln Arg Ser Pro Ser Asp Ala Gly Lys Ser Ser Gly Asp	2260
gaa ggg aaa aag ccc ccc tca ggc att gga aga tcg act gcc acc agc Glu Gly Lys Lys Pro Pro Ser Gly Ile Gly Arg Ser Thr Ala Thr Ser	2308
tcc ttt ggc ttt aag aaa cca agt gga gta ggg tca tct gcc atg atc Ser Phe Gly Phe Lys Lys Pro Ser Gly Val Gly Ser Ser Ala Met Ile	2356
acc agc agt gga gca acc ata aca agt ggc tct gca aca ctg ggt aaa Thr Ser Ser Gly Ala Thr Ile Thr Ser Gly Ser Ala Thr Leu Gly Lys	2404
att cca aaa tct gct gcc att ggc ggg aag tca aat gca ggg aga aaa Ile Pro Lys Ser Ala Ala Ile Gly Gly Lys Ser Asn Ala Gly Arg Lys	2452
acc agt ttg gac ggt tca cag aat cag gat gat gtt gtg ctg cat gtt Thr Ser Leu Asp Gly Ser Gln Asn Gln Asp Asp Val Val Leu His Val	2500
agc tca aag act acc cta caa tat cgc agc ttg ccc cgc cct tca aaa Ser Ser Lys Thr Thr Leu Gln Tyr Arg Ser Leu Pro Arg Pro Ser Lys	2548
tcc agc acc agt ggc att cct ggc cga gga ggc cac aga tcc agt acc Ser Ser Thr Ser Gly Ile Pro Gly Arg Gly Gly His Arg Ser Ser Thr	2596
agc agt att gat tcc aac gtc agc agc aag tct gct ggg gcc acc acc Ser Ser Ile Asp Ser Asn Val Ser Ser Lys Ser Ala Gly Ala Thr Thr	2644
tcg aaa ctg aga gaa cca act aaa att ggg tca ggg cgc tcg agt cct Ser Lys Leu Arg Glu Pro Thr Lys Ile Gly Ser Gly Arg Ser Ser Pro	2692
gtc acc gtc aac caa aca gac aag gaa aag gaa aaa gta gca gtc tca Val Thr Val Asn Gln Thr Asp Lys Glu Lys Glu Lys Val Ala Val Ser	2740
gat tca gaa agt gtt tct ttg tca ggt tcc ccc aaa tcc agc ccc acc Asp Ser Glu Ser Val Ser Leu Ser Gly Ser Pro Lys Ser Ser Pro Thr	2788

tct gcc agc gcc tgt ggt gca caa ggt ctc agg cag cca gga tcc aag Ser Ala Ser Ala Cys Gly Ala Gln Gly Leu Arg Gln Pro Gly Ser Lys	2836
tat cca gat att gcc tca ccc aca ttt cga agg ttg ttt ggt gcc aag Tyr Pro Asp Ile Ala Ser Pro Thr Phe Arg Arg Leu Phe Gly Ala Lys	2884
gca ggt ggc aaa tct gcc tct gca cct aat act gag ggt gtg aaa tct Ala Gly Gly Lys Ser Ala Ser Ala Pro Asn Thr Glu Gly Val Lys Ser	2932
tcc tca gta atg ccc agc cct agt acc aca tta gcg cgg caa ggc agt Ser Ser Val Met Pro Ser Pro Ser Thr Thr Leu Ala Arg Gln Gly Ser	2980
ctg gag tca ccg tcg tcc ggt acg ggc agc atg ggc agt gct ggt ggg Leu Glu Ser Pro Ser Ser Gly Thr Gly Ser Met Gly Ser Ala Gly Gly	3028
cta agc ggc agc agc agc cct ctc ttc aat aaa ccc tca gac tta act Leu Ser Gly Ser Ser Ser Pro Leu Phe Asn Lys Pro Ser Asp Leu Thr	3076
aca gat gtt ata agc tta agt cac tcg ttg gcc tcc agc cca gca tcg Thr Asp Val Ile Ser Leu Ser His Ser Leu Ala Ser Ser Pro Ala Ser	3124
gtt cac tct ttc aca tca ggt ggt ctc gtg tgg gct gcc aat atg agc Val His Ser Phe Thr Ser Gly Gly Leu Val Trp Ala Ala Asn Met Ser	3172
agt tcc tct gca ggc agc aag gat act ccg agc tac cag tcc atg act Ser Ser Ser Ala Gly Ser Lys Asp Thr Pro Ser Tyr Gln Ser Met Thr	3220
agc ctc cac acg agc tct gag tcc att gac ctc ccc ctc agc cat cat Ser Leu His Thr Ser Ser Glu Ser Ile Asp Leu Pro Leu Ser His His	3268
ggc tcc ttg tct gga ctg acc aca ggc act cac gag gtc cag agc ctg Gly Ser Leu Ser Gly Leu Thr Thr Gly Thr His Glu Val Gln Ser Leu	3316
ctc atg aga acg ggt agt gtg aga tct act ctc tca gaa agc atg cag Leu Met Arg Thr Gly Ser Val Arg Ser Thr Leu Ser Glu Ser Met Gln	3364
ctt gac aga aat aca cta ccc aaa aag gga cta aga tat acc cca tca Leu Asp Arg Asn Thr Leu Pro Lys Lys Gly Leu Arg Tyr Thr Pro Ser	3412
tct cgg cag gcc aac caa gaa gag ggc aaa gag tgg ttg cgt tct cat Ser Arg Gln Ala Asn Gln Glu Glu Gly Lys Glu Trp Leu Arg Ser His	3460
tct act gga ggg ctt cag gac act ggc aac cag tca cct ctg gtt tcc	3508

Ser Thr Gly Gly Leu Gln Asp Thr Gly Asn Gln Ser Pro Leu Val Ser

cct tct gcc atg tca tct tct gca gct gga aaa tac cac ttt tct aac 3556  
Pro Ser Ala Met Ser Ser Ser Ala Ala Gly Lys Tyr His Phe Ser Asn

ttg gtg agc cca aca aat ttg tct caa ttt aac ctt ccc ggg ccc agc 3604  
Leu Val Ser Pro Thr Asn Leu Ser Gln Phe Asn Leu Pro Gly Pro Ser

atg atg cgc tca aac agc atc cca gcc caa gac tct tcc ttc gat ctc 3652  
Met Met Arg Ser Asn Ser Ile Pro Ala Gln Asp Ser Ser Phe Asp Leu

tat gat gac tcc cag ctt tgt ggg agt gcc act tct ctg gag gaa aga 3700  
Tyr Asp Asp Ser Gln Leu Cys Gly Ser Ala Thr Ser Leu Glu Glu Arg

cct cgt gcc atc agt cat tcg ggc tca ttc aga gac agc atg gaa gaa 3748  
Pro Arg Ala Ile Ser His Ser Gly Ser Phe Arg Asp Ser Met Glu Glu

gtt cat ggc tct tca tta tca ctg gtg tcc agc act tct tct ctt tac 3796  
Val His Gly Ser Ser Leu Ser Leu Val Ser Ser Thr Ser Ser Leu Tyr

tct aca gct gaa gaa aag gct cat tca gag caa atc cat aaa ctg cgg 3844  
Ser Thr Ala Glu Glu Lys Ala His Ser Glu Gln Ile His Lys Leu Arg

aga gag ctg gtt gca tca caa gaa aaa gtt gct acc ctc aca tct cag 3892  
Arg Glu Leu Val Ala Ser Gln Glu Lys Val Ala Thr Leu Thr Ser Gln

ctt tca gca aat gct cac ctt gta gca gct ttt gaa aag agc tta ggg 3940  
Leu Ser Ala Asn Ala His Leu Val Ala Ala Phe Glu Lys Ser Leu Gly

aat atg act ggc cga ttg caa agt cta act atg aca gcg gaa caa aag 3988  
Asn Met Thr Gly Arg Leu Gln Ser Leu Thr Met Thr Ala Glu Gln Lys

gaa tct gaa ctt ata gaa cta aga gaa acc att gaa atg ctg aag gct 4036  
Glu Ser Glu Leu Ile Glu Leu Arg Glu Thr Ile Glu Met Leu Lys Ala

cag aat tct gct gcc cag gcg gct att cag gga gca ctg aat ggt cca 4084  
Gln Asn Ser Ala Ala Gln Ala Ala Ile Gln Gly Ala Leu Asn Gly Pro

gac cat cct ccc aaa gat ctt cgc atc aga aga cag cat tcc tct gaa 4132  
Asp His Pro Pro Lys Asp Leu Arg Ile Arg Arg Gln His Ser Ser Glu

agt gtt tct agt atc aac agt gcc aca agc cat tcc agt att ggc agt 4180  
Ser Val Ser Ser Ile Asn Ser Ala Thr Ser His Ser Ser Ile Gly Ser

ggt aat gat gcc gac tcc aag aag aag aaa aag aaa aac tgg ctg aga Gly Asn Asp Ala Asp Ser Lys Lys Lys Lys Lys Lys Asn Trp Leu Arg	4228
agt tct ttc aaa caa gcc ttt ggg aag aaa aag tcc acc aag cct cct Ser Ser Phe Lys Gln Ala Phe Gly Lys Lys Lys Ser Thr Lys Pro Pro	4276
tca tca cat tct gac att gaa gag ctt act gat tca tcc ctt ccg gca Ser Ser His Ser Asp Ile Glu Glu Leu Thr Asp Ser Ser Leu Pro Ala	4324
tcc ccc aag tta ccc cat aat gct ggt gac tgt ggc tca gca tcc atg Ser Pro Lys Leu Pro His Asn Ala Gly Asp Cys Gly Ser Ala Ser Met	4372
aag ccc tca caa tct gct tca gcg atc tgt gaa tgc aca gaa gct gag Lys Pro Ser Gln Ser Ala Ser Ala Ile Cys Glu Cys Thr Glu Ala Glu	4420
gca gag ata att ctg cag ctg aag agc gag ctc aga gaa aag gaa tta Ala Glu Ile Ile Leu Gln Leu Lys Ser Glu Leu Arg Glu Lys Glu Leu	4468
aaa tta acg gat att cgg ctg gag gcc ctc agc tct gct cat cat ctt Lys Leu Thr Asp Ile Arg Leu Glu Ala Leu Ser Ser Ala His His Leu	4516
gat cag atc cgg gaa gcc atg aac cgg atg cag aat gaa att gaa ata Asp Gln Ile Arg Glu Ala Met Asn Arg Met Gln Asn Glu Ile Glu Ile	4564
ctg aaa gct gaa aat gac cgg ttg aag gca gaa act ggt aac aca gct Leu Lys Ala Glu Asn Asp Arg Leu Lys Ala Glu Thr Gly Asn Thr Ala	4612
aag cct act cgg cca ccg tca gaa tcc tca agc agc acc tcc tct tca Lys Pro Thr Arg Pro Pro Ser Glu Ser Ser Ser Ser Thr Ser Ser Ser	4660
tct tcc agg cag tca tta gga ctt tct cta aac aat ttg aac atc aca Ser Ser Arg Gln Ser Leu Gly Leu Ser Leu Asn Asn Leu Asn Ile Thr	4708
gag gct gtt agc tca gat att ttg cta gat gat gct ggt gat gca act Glu Ala Val Ser Ser Asp Ile Leu Leu Asp Asp Ala Gly Asp Ala Thr	4756
gga cat aaa gat ggc cgc agt gtg aaa att ata gtc tcc ata agc aag Gly His Lys Asp Gly Arg Ser Val Lys Ile Ile Val Ser Ile Ser Lys	4804
ggc tat ggt cga gca aag gac caa aaa tct cag gca tat ttg ata gga Gly Tyr Gly Arg Ala Lys Asp Gln Lys Ser Gln Ala Tyr Leu Ile Gly	4852

tcc att ggt gtt agt gga aaa acc aag tgg gat gtc tta gat ggt gta Ser Ile Gly Val Ser Gly Lys Thr Lys Trp Asp Val Leu Asp Gly Val	4900
ata aga cgt ctc ttt aag gaa tat gta ttc cga att gat aca tcc act Ile Arg Arg Leu Phe Lys Glu Tyr Val Phe Arg Ile Asp Thr Ser Thr	4948
agc ctt ggt ctg agc tct gac tgc att gct agc tac tgt ata gga gac Ser Leu Gly Leu Ser Ser Asp Cys Ile Ala Ser Tyr Cys Ile Gly Asp	4996
tta att aga tcc cat aac cta gaa gtg cct gaa ttg ctg cct tgt gga Leu Ile Arg Ser His Asn Leu Glu Val Pro Glu Leu Leu Pro Cys Gly	5044
tac ctt gtt gga gat aat aac atc atc act gtg aac ctc aaa ggg gta Tyr Leu Val Gly Asp Asn Asn Ile Ile Thr Val Asn Leu Lys Gly Val	5092
gaa gaa aat agt ttg gac agt ttt gtt ttt gat acg ctg att cct aaa Glu Glu Asn Ser Leu Asp Ser Phe Val Phe Asp Thr Leu Ile Pro Lys	5140
cca att acc caa agg tac ttt aac ttg ttg atg gag cat cac aga att Pro Ile Thr Gln Arg Tyr Phe Asn Leu Leu Met Glu His His Arg Ile	5188
ata ctc tca gga ccg agt ggt act gga aag acc tat ttg gca aac aaa Ile Leu Ser Gly Pro Ser Gly Thr Gly Lys Thr Tyr Leu Ala Asn Lys	5236
ctt gct gaa tat gta ata acc aaa tct gga agg aaa aaa aca gag gat Leu Ala Glu Tyr Val Ile Thr Lys Ser Gly Arg Lys Lys Thr Glu Asp	5284
gca att gcc act ttt aat gtg gac cac aag tca agt aag gaa ttg caa Ala Ile Ala Thr Phe Asn Val Asp His Lys Ser Ser Lys Glu Leu Gln	5332
caa tat cta gct aac ctg gct gaa cag tgc agt gct gat aat aat gga Gln Tyr Leu Ala Asn Leu Ala Glu Gln Cys Ser Ala Asp Asn Asn Gly	5380
gtg gag ctc cca gtt gta ata att ctt gat aat ctt cat cat gtg ggc Val Glu Leu Pro Val Val Ile Ile Leu Asp Asn Leu His His Val Gly	5428
tct ctg agt gat atc ttc aat ggt ttt ctc aat tgt aaa tac aac aaa Ser Leu Ser Asp Ile Phe Asn Gly Phe Leu Asn Cys Lys Tyr Asn Lys	5476
tgt cca tat att att gga aca atg aat cag gga gtt tct tca tca cca Cys Pro Tyr Ile Ile Gly Thr Met Asn Gln Gly Val Ser Ser Ser Pro	5524

aat cta gag ctg cat cac aat ttc agg tgg gta tta tgt gca aat cat Asn Leu Glu Leu His His Asn Phe Arg Trp Val Leu Cys Ala Asn His	5572
aca gaa cca gtg aaa ggc ttt tta ggc aga tat ctt cga aga aaa ctc Thr Glu Pro Val Lys Gly Phe Leu Gly Arg Tyr Leu Arg Arg Lys Leu	5620
ata gag ata gaa att gaa agg aac att cgc aat aat gac cta gtc aaa Ile Glu Ile Glu Ile Glu Arg Asn Ile Arg Asn Asn Asp Leu Val Lys	5668
att ata gat tgg att ccg aag acg tgg cat cat ctc aac agt ttt ttg Ile Ile Asp Trp Ile Pro Lys Thr Trp His His Leu Asn Ser Phe Leu	5716
gaa aca cac agt tct tct gac gtt acc att ggt ccc cga cta ttc ctt Glu Thr His Ser Ser Ser Asp Val Thr Ile Gly Pro Arg Leu Phe Leu	5764
cct tgc ccc atg gat gta gaa ggt tct aga gta tgg ttc atg gat ctc Pro Cys Pro Met Asp Val Glu Gly Ser Arg Val Trp Phe Met Asp Leu	5812
tgg aac tat tct tta gta cct tat att ctg gag gca gtg aga gag ggt Trp Asn Tyr Ser Leu Val Pro Tyr Ile Leu Glu Ala Val Arg Glu Gly	5860
ctt cag atg tat ggg aaa cgc aca cca tgg gaa gat cct tca aag tgg Leu Gln Met Tyr Gly Lys Arg Thr Pro Trp Glu Asp Pro Ser Lys Trp	5908
gtg ctt gac aca tat cca tgg agc tca gca act ctg cct cag gag agc Val Leu Asp Thr Tyr Pro Trp Ser Ser Ala Thr Leu Pro Gln Glu Ser	5956
cca gcc tta ctt cag ctg cga cca gaa gat gtt ggg tat gaa agc tgc Pro Ala Leu Leu Gln Leu Arg Pro Glu Asp Val Gly Tyr Glu Ser Cys	6004
aca tcc act aag gaa gcc aca acc tca aag cac att cca caa act gac Thr Ser Thr Lys Glu Ala Thr Thr Ser Lys His Ile Pro Gln Thr Asp	6052
aca gaa gga gat ccc ctg atg aat atg cta atg aaa ctc caa gaa gca Thr Glu Gly Asp Pro Leu Met Asn Met Leu Met Lys Leu Gln Glu Ala	6100
gcc aat tac tcg agc aca caa agc tgc gac agc gaa agc acc agt cac Ala Asn Tyr Ser Ser Thr Gln Ser Cys Asp Ser Glu Ser Thr Ser His	6148
cat gaa gac att ttg gat tca tct ctt gaa tct acc ctc tagaggggtga His Glu Asp Ile Leu Asp Ser Ser Leu Glu Ser Thr Leu	6197
aaaaagttaa gggaaaagac tttgctttta aaaaaatggt tcaaaagaaa ggtattttca	6257

ctaaaccact gccagtataa aagcacccctg tcaagggccc tgaccagag ttgtggtctc	6317
caaggaggca gcagaactaa gtctgaaccg ccaagatgct aaattgcaat ggaagcttaa	6377
ctttagttta tttctaaaca ttttttatat ctgtggagta atagaaagct ccattactca	6437
actggaaagg accctaataa cagggcaact gaacagattg cacatgggat agccaaactg	6497
gactttcttt gtttctcttt taaaagttaa caatgcagac cttttttgt cccttcttt	6557
tgtttctctt gaggggctgt tggccccagg caggggtccat ctttctgac tgtccaacct	6617
cctttgtgcc acacggtgct ggtcacaggg cttcagtagt gtttgtgttg tgcgctcacc	6677
ccattccaga acaaatacaa gaggccagtc ctccataagc acaaataaggaa ttgtgcaacc	6737
accagaaaaa cactactgtg gcaaaactgga gaagtgccaa ttttaattcta actgccacgt	6797
tctcatgatg tgctccacca acttttttagt atatgagtca ctggttttat aagggtgttt	6857
ttaccacagt ggtcttttta aaccacctgc cactccctt aacaagagtt ttataccaat	6917
tattagtcaa cactgataaa aggcctttttt agggctttat ttgtttgagc cttttcagtg	6977
aaagaaggaa catttcctat ggtgctgtct cactgcctta aaacagattt ctatgacagt	7037
ttaacagttg gtttaaatcc taaaccattg gtaatttcca ctgtcttttc atttacaacc	7097
aagcaacacc agttaacata gtagcctcat ctctatatat ctttctcttt tttttttttt	7157
tgaagaaatg gataggagaa agatcagtat ttttagcctt gtgaatagat cgctttgcct	7217
atcctccaaa atattaaaat aaccagaaa tgctctttga ccgtcactta aaacctaaga	7277
catgtggcga aattccatcc agttctaagt gaaagagttt cagaaggcag gagattttga	7337
attattatcc agcagggtg gaagcactag atgcagcatg agcacacta ttcggctttc	7397
cttccctatt gttttgttt ttttaatgag ttttgacgca tgttgttttg attgctattg	7457
ttgtacatga gaaattcagc attaaagaac actgaagcgg taaggctact gtggaagagg	7517
aagcgtttat actgtaaaag aaggttagat ttgcacagtc tactgggtag gtattgtaaa	7577
taataatttt taaaacttgc acaaatcaaa acaaacacaa acaaaattgt attttatcct	7637
gttggtgtta agagggtgtt cacttgctga gatttctgt acattgcaaa caaatacaga	7697
atgcaaacc tcaaagctgt attatctggt gtgtttgtcc tgtatttaca gttgttttg	7757
actatgcagg agctatcagt gctagagtga gcatgcttca aaactgtaca tgaagccaat	7817
atatttttgg ataagtaaaa ctgtctgaaa gtacatctgt catggcaggc tttaaagaga	7877
gtgcatgaaa actgatcagt cattggagaa gttaccacca cacacaaagg acaggtttta	7937



agtttatgaa acccaagggc taggcatgg tatagacttc ttctatgagt gtgtgaaaat 7997  
 gtgttacttt taggacgtgt atttgggtgct actctctgtg accaccaatg ggtcagttgc 8057  
 tatagaacaa caacaccacg aaacatctgt gcagttttca gagtgtcaca aagtcaatag 8117  
 gtccttacac ggtgctattg ccctaagggg aatccgaact gaatttatgc acatagaatt 8177  
 gtcaccctga ctttgaagcc tcaaacatgg atcaaatctg ttgtgaaaca tcaatatatg 8237  
 tagctggatg agtgactagt ttcccttgta taatatgtga tctaagaaaa ttgctaattc 8297  
 ttccctgcca ttttgagaaa cacagtccaa acatgagcat aaacagaatt tcctgcaata 8357  
 catcccagta ggtccaccta gtttacaact taaactagtt tgtgaaacat ttgtctgtat 8417  
 acattttata ttttgtacat ttgatgtaa catatcatgt aaataggcag aaacagtgaa 8477  
 ataaatcatc tgaaaagttt tgtagtcttt gttaaagcccc aacaataagt acttgggtgc 8537  
 aatggactta actggatgat gtattttcta ttggtttatt gttcctctag cttgtaaacc 8597  
 agcttgcata tatttttttg caaatgtgca ccctgtatct gtctaaatta ttactttgcc 8657  
 attaaagtgg aattatttat tgacaaaaaa aaa 8690

<210> 12  
 <211> 1895  
 <212> PRT  
 <213> Homo sapiens

<400> 12

Met Asp Leu Ser Ser Glu Met Asn Arg His Gly Lys Asn Pro Val Ser  
 1 5 10 15  
 His Lys Leu Glu Asp Gln Lys Lys Ile Tyr Thr Asp Trp Ala Asn His  
 20 25 30  
 Tyr Leu Ala Lys Ser Gly His Lys Arg Leu Ile Lys Asp Leu Gln Gln  
 35 40 45  
 Asp Ile Ala Asp Gly Val Leu Leu Ala Glu Ile Ile Gln Ile Ile Ala  
 50 55 60  
 Asn Glu Lys Val Glu Asp Ile Asn Gly Cys Pro Arg Ser Gln Ser Gln  
 65 70 75 80  
 Met Ile Glu Asn Val Asp Val Cys Leu Ser Phe Leu Ala Ala Arg Gly  
 85 90 95  
 Val Asn Val Gln Gly Leu Ser Ala Glu Glu Ile Arg Asn Gly Asn Leu  
 100 105 110  
 Lys Ala Ile Leu Gly Leu Phe Phe Ser Leu Ser Arg Tyr Lys Gln Gln

115				120				125								
Gln	His	His	Gln	Gln	Gln	Tyr	Tyr	Gln	Ser	Leu	Val	Glu	Leu	Gln	Gln	
130						135						140				
Arg	Val	Thr	His	Ala	Ser	Pro	Pro	Ser	Glu	Ala	Ser	Gln	Ala	Lys	Thr	
145					150			155			160					
Gln	Gln	Asp	Met	Gln	Ser	Arg	Leu	Pro	Gly	Pro	Ser	Arg	Val	Pro	Ala	
				165			170						175			
Ala	Gly	Ser	Ser	Ser	Lys	Val	Gln	Gly	Ala	Ser	Asn	Leu	Asn	Arg	Arg	
			180			185						190				
Ser	Gln	Ser	Phe	Asn	Ser	Ile	Asp	Lys	Asn	Lys	Pro	Pro	Asn	Tyr	Ala	
		195			200			205								
Asn	Gly	Asn	Glu	Lys	Gly	Glu	Asp	Pro	Glu	Thr	Arg	Arg	Met	Arg	Thr	
210					215			220								
Val	Lys	Asn	Ile	Ala	Asp	Leu	Arg	Gln	Asn	Leu	Glu	Glu	Thr	Met	Ser	
225					230			235			240					
Ser	Leu	Arg	Gly	Thr	Gln	Ile	Ser	His	Ser	Thr	Leu	Glu	Thr	Thr	Phe	
				245			250						255			
Asp	Ser	Thr	Val	Thr	Thr	Glu	Val	Asn	Gly	Arg	Thr	Ile	Pro	Asn	Leu	
			260			265						270				
Thr	Ser	Arg	Pro	Thr	Pro	Met	Thr	Trp	Arg	Leu	Gly	Gln	Ala	Cys	Pro	
		275			280						285					
Arg	Leu	Gln	Ala	Gly	Asp	Ala	Pro	Ser	Leu	Gly	Ala	Gly	Tyr	Pro	Arg	
290					295			300								
Ser	Gly	Thr	Ser	Arg	Phe	Ile	His	Thr	Asp	Pro	Ser	Arg	Phe	Met	Tyr	
305					310			315			320					
Thr	Thr	Pro	Leu	Arg	Arg	Ala	Ala	Val	Ser	Arg	Leu	Gly	Asn	Met	Ser	
			325			330						335				
Gln	Ile	Asp	Met	Ser	Glu	Lys	Ala	Ser	Ser	Asp	Leu	Asp	Met	Ser	Ser	
			340			345						350				
Glu	Val	Asp	Val	Gly	Gly	Tyr	Met	Ser	Asp	Gly	Asp	Ile	Leu	Gly	Lys	
355					360			365								
Ser	Leu	Arg	Thr	Asp	Asp	Ile	Asn	Ser	Gly	Tyr	Met	Thr	Asp	Gly	Gly	
370					375			380								
Leu	Asn	Leu	Tyr	Thr	Arg	Ser	Leu	Asn	Arg	Ile	Pro	Asp	Thr	Ala	Thr	
385					390			395			400					
Ser	Arg	Asp	Ile	Ile	Gln	Arg	Gly	Val	His	Asp	Val	Thr	Val	Asp	Ala	
			405			410						415				
Asp	Ser	Trp	Asp	Asp	Ser	Ser	Ser	Val	Ser	Ser	Gly	Leu	Ser	Asp	Thr	



Thr	Lys	Ile	Gly	Ser	Gly	Arg	Ser	Pro	Val	Thr	Val	Asn	Gln	Thr	
				725					730					735	
Asp	Lys	Glu	Lys	Glu	Lys	Val	Ala	Val	Ser	Asp	Ser	Glu	Ser	Val	Ser
			740					745					750		
Leu	Ser	Gly	Ser	Pro	Lys	Ser	Ser	Pro	Thr	Ser	Ala	Ser	Ala	Cys	Gly
		755					760					765			
Ala	Gln	Gly	Leu	Arg	Gln	Pro	Gly	Ser	Lys	Tyr	Pro	Asp	Ile	Ala	Ser
	770					775					780				
Pro	Thr	Phe	Arg	Arg	Leu	Phe	Gly	Ala	Lys	Ala	Gly	Gly	Lys	Ser	Ala
785					790					795					800
Ser	Ala	Pro	Asn	Thr	Glu	Gly	Val	Lys	Ser	Ser	Ser	Val	Met	Pro	Ser
				805					810					815	
Pro	Ser	Thr	Thr	Leu	Ala	Arg	Gln	Gly	Ser	Leu	Glu	Ser	Pro	Ser	Ser
			820					825					830		
Gly	Thr	Gly	Ser	Met	Gly	Ser	Ala	Gly	Gly	Leu	Ser	Gly	Ser	Ser	Ser
		835					840					845			
Pro	Leu	Phe	Asn	Lys	Pro	Ser	Asp	Leu	Thr	Thr	Asp	Val	Ile	Ser	Leu
	850					855					860				
Ser	His	Ser	Leu	Ala	Ser	Ser	Pro	Ala	Ser	Val	His	Ser	Phe	Thr	Ser
865					870					875					880
Gly	Gly	Leu	Val	Trp	Ala	Ala	Asn	Met	Ser	Ser	Ser	Ser	Ala	Gly	Ser
				885					890					895	
Lys	Asp	Thr	Pro	Ser	Tyr	Gln	Ser	Met	Thr	Ser	Leu	His	Thr	Ser	Ser
			900					905					910		
Glu	Ser	Ile	Asp	Leu	Pro	Leu	Ser	His	His	Gly	Ser	Leu	Ser	Gly	Leu
		915					920					925			
Thr	Thr	Gly	Thr	His	Glu	Val	Gln	Ser	Leu	Leu	Met	Arg	Thr	Gly	Ser
	930					935					940				
Val	Arg	Ser	Thr	Leu	Ser	Glu	Ser	Met	Gln	Leu	Asp	Arg	Asn	Thr	Leu
945					950					955					960
Pro	Lys	Lys	Gly	Leu	Arg	Tyr	Thr	Pro	Ser	Ser	Arg	Gln	Ala	Asn	Gln
				965					970					975	
Glu	Glu	Gly	Lys	Glu	Trp	Leu	Arg	Ser	His	Ser	Thr	Gly	Gly	Leu	Gln
			980					985					990		
Asp	Thr	Gly	Asn	Gln	Ser	Pro	Leu	Val	Ser	Pro	Ser	Ala	Met	Ser	Ser
		995					1000					1005			
Ser	Ala	Ala	Gly	Lys	Tyr	His	Phe	Ser	Asn	Leu	Val	Ser	Pro	Thr	Asn
	1010					1015					1020				

Leu Ser Gln Phe Asn Leu Pro Gly Pro Ser Met Met Arg Ser Asn Ser  
 1025 1030 1035 1040  
 Ile Pro Ala Gln Asp Ser Ser Phe Asp Leu Tyr Asp Asp Ser Gln Leu  
 1045 1050 1055  
 Cys Gly Ser Ala Thr Ser Leu Glu Glu Arg Pro Arg Ala Ile Ser His  
 1060 1065 1070  
 Ser Gly Ser Phe Arg Asp Ser Met Glu Glu Val His Gly Ser Ser Leu  
 1075 1080 1085  
 Ser Leu Val Ser Ser Thr Ser Ser Leu Tyr Ser Thr Ala Glu Glu Lys  
 1090 1095 1100  
 Ala His Ser Glu Gln Ile His Lys Leu Arg Arg Glu Leu Val Ala Ser  
 1105 1110 1115 1120  
 Gln Glu Lys Val Ala Thr Leu Thr Ser Gln Leu Ser Ala Asn Ala His  
 1125 1130 1135  
 Leu Val Ala Ala Phe Glu Lys Ser Leu Gly Asn Met Thr Gly Arg Leu  
 1140 1145 1150  
 Gln Ser Leu Thr Met Thr Ala Glu Gln Lys Glu Ser Glu Leu Ile Glu  
 1155 1160 1165  
 Leu Arg Glu Thr Ile Glu Met Leu Lys Ala Gln Asn Ser Ala Ala Gln  
 1170 1175 1180  
 Ala Ala Ile Gln Gly Ala Leu Asn Gly Pro Asp His Pro Pro Lys Asp  
 1185 1190 1195 1200  
 Leu Arg Ile Arg Arg Gln His Ser Ser Glu Ser Val Ser Ser Ile Asn  
 1205 1210 1215  
 Ser Ala Thr Ser His Ser Ser Ile Gly Ser Gly Asn Asp Ala Asp Ser  
 1220 1225 1230  
 Lys Lys Lys Lys Lys Lys Asn Trp Leu Arg Ser Ser Phe Lys Gln Ala  
 1235 1240 1245  
 Phe Gly Lys Lys Lys Ser Thr Lys Pro Pro Ser Ser His Ser Asp Ile  
 1250 1255 1260  
 Glu Glu Leu Thr Asp Ser Ser Leu Pro Ala Ser Pro Lys Leu Pro His  
 1265 1270 1275 1280  
 Asn Ala Gly Asp Cys Gly Ser Ala Ser Met Lys Pro Ser Gln Ser Ala  
 1285 1290 1295  
 Ser Ala Ile Cys Glu Cys Thr Glu Ala Glu Ala Glu Ile Ile Leu Gln  
 1300 1305 1310  
 Leu Lys Ser Glu Leu Arg Glu Lys Glu Leu Lys Leu Thr Asp Ile Arg

1315					1320					1325						
Leu	Glu	Ala	Leu	Ser	Ser	Ala	His	His	Leu	Asp	Gln	Ile	Arg	Glu	Ala	
1330					1335					1340						
Met	Asn	Arg	Met	Gln	Asn	Glu	Ile	Glu	Ile	Leu	Lys	Ala	Glu	Asn	Asp	
1345					1350					1355					1360	
Arg	Leu	Lys	Ala	Glu	Thr	Gly	Asn	Thr	Ala	Lys	Pro	Thr	Arg	Pro	Pro	
1365					1370					1375						
Ser	Glu	Ser	Ser	Ser	Ser	Thr	Ser	Ser	Ser	Ser	Ser	Arg	Gln	Ser	Leu	
1380					1385					1390						
Gly	Leu	Ser	Leu	Asn	Asn	Leu	Asn	Ile	Thr	Glu	Ala	Val	Ser	Ser	Asp	
1395					1400					1405						
Ile	Leu	Leu	Asp	Asp	Ala	Gly	Asp	Ala	Thr	Gly	His	Lys	Asp	Gly	Arg	
1410					1415					1420						
Ser	Val	Lys	Ile	Ile	Val	Ser	Ile	Ser	Lys	Gly	Tyr	Gly	Arg	Ala	Lys	
1425					1430					1435					1440	
Asp	Gln	Lys	Ser	Gln	Ala	Tyr	Leu	Ile	Gly	Ser	Ile	Gly	Val	Ser	Gly	
1445					1450					1455						
Lys	Thr	Lys	Trp	Asp	Val	Leu	Asp	Gly	Val	Ile	Arg	Arg	Leu	Phe	Lys	
1460					1465					1470						
Glu	Tyr	Val	Phe	Arg	Ile	Asp	Thr	Ser	Thr	Ser	Leu	Gly	Leu	Ser	Ser	
1475					1480					1485						
Asp	Cys	Ile	Ala	Ser	Tyr	Cys	Ile	Gly	Asp	Leu	Ile	Arg	Ser	His	Asn	
1490					1495					1500						
Leu	Glu	Val	Pro	Glu	Leu	Leu	Pro	Cys	Gly	Tyr	Leu	Val	Gly	Asp	Asn	
1505					1510					1515					1520	
Asn	Ile	Ile	Thr	Val	Asn	Leu	Lys	Gly	Val	Glu	Glu	Asn	Ser	Leu	Asp	
1525					1530					1535						
Ser	Phe	Val	Phe	Asp	Thr	Leu	Ile	Pro	Lys	Pro	Ile	Thr	Gln	Arg	Tyr	
1540					1545					1550						
Phe	Asn	Leu	Leu	Met	Glu	His	His	Arg	Ile	Ile	Leu	Ser	Gly	Pro	Ser	
1555					1560					1565						
Gly	Thr	Gly	Lys	Thr	Tyr	Leu	Ala	Asn	Lys	Leu	Ala	Glu	Tyr	Val	Ile	
1570					1575					1580						
Thr	Lys	Ser	Gly	Arg	Lys	Lys	Thr	Glu	Asp	Ala	Ile	Ala	Thr	Phe	Asn	
1585					1590					1595					1600	
Val	Asp	His	Lys	Ser	Ser	Lys	Glu	Leu	Gln	Gln	Tyr	Leu	Ala	Asn	Leu	
1605					1610					1615						



<211> 3682  
 <212> DNA  
 <213> Homo sapiens

<400> 13

gaattcctgg tggagaacag cacatgtaca gatgggggtga gaacagcata cgtacaggta	60
ggggtaagct ggtgctatat gagaaagcat ggaataagtt attaagtttg acctgcttgg	120
gaactgaggg gcagggtgtga gggatgaagc aggagtaggt aggggctaga tcacaaaaga	180
tctatgccag tgtttctcac agtgtgattc ccagcccagt agcatgatata cacttgggat	240
cttggttagaa atacaaattc ttatacatca ccttggacta gaccacctga ataagaaaag	300
ttgggcatga ggcctacaaa tttttaaaaa agtcatacag gtgattgcaa tgcattgctaa	360
agtttgagaa acactctttg ctgtgggttg aatatttgtg tccttccaaa attcatgtag	420
aaaccatctc caatgttata gtattaagag gagggacct tgggagctga tcagatcatg	480
aagtctcctt tcttataaag gggattaaaa gccttggccc ttttacctt tgtccatgta	540
aggacacagt gttggaagca gggactgggt tctcaccaga aacagaacct gccagcctct	600
tggctcttga cttctcagcc tccacaattg tgagaaataa gtttctgttg tttataagtt	660
aaccagtctc aggtattttg taatggcagc acaaaggggc taagaaactg ttctatgccc	720
taacaagaaa tgtggtcact ttctgaagg aaatggggat atatataaag atgttatata	780
agactcgtaa tatttatttg gaaggcttgc tctgcaagca aggtggaaga gcaacatgaa	840
ggaagcgtgg tggaggtgag aggactggag gttaagttgg tagggagata caggaaagaa	900
gcttatgaca cttgagttaa aatgtagcat ccttcctatg tgtagggctc ataaaaatgt	960
atagtctaag atagaacaca gaatactcta tgaatcctgc ccacaagggtg ttggtaatct	1020
agattcactt ttttttctg ataatgccat ccatatgtat ggagcgtcta ctactgtatg	1080
ccagagtga ccttggaaatcg gtttggttga tctagacaag accataagga gactcccctt	1140
actacctctt ctccagggga gggattcaag ttgaactagt acttcagaga ctgttttagta	1200
atatcatgca tgaaagggtga tggttaggac agaaaaataa atggattgca tcataattcc	1260
tcaggttctc caaatatgtg gtggtctcaa accatgtgaa ttggtctgca catcctgttt	1320
ggggttgcgtg tcagcagttg agatctgagc cttatttgta acagtgaaac agtgagagac	1380
ctgcccttca agagctgttt tttagctagg aatagaaaag ggccaggcta gactcctctt	1440
tctgctggat cttgcttctt ctcagcaata gaagtagacc tgccttccta gctgtagaga	1500



aaaggtgccc	gtaggcgggc	aggtgagcct	gtggataatc	ctggagtaaa	ggttcaatag	1560
accttcaagt	ctatcctaca	ggattcggag	tgaggggaga	gaaaaggaga	cgttctctcg	1620
gctgagagag	gaagagaaaa	aaaaatccca	gatatctgac	agctatatct	tcccatcacc	1680
accttctctc	aaacccatgc	ctctctgttt	agtaggacat	aaaatgaaga	gtgacccacc	1740
ccccaccccc	agcccatccc	ccgtttgtag	gtgtgctttc	aatgaaaata	agtcgggtgt	1800
catggacgga	aactagagca	gctgaaaata	gatgcaagac	ttgttgagca	tacaaatcat	1860
ttccccctta	gtctccaagg	gaggaaaaaa	aatccctctt	actctccttg	cagcctgtgt	1920
tctgcattct	ggagaggaa	ctgaggctgg	tctcaggcg	ctcctcccg	cgttcccgca	1980
ggaaactttt	ctcgcagggc	ccgctccgtc	catcccgcgc	ggttccaaga	cggtgggcct	2040
cccggtgggct	cctctcctgg	gcaagggccc	agaccccgcg	acgcgcctgt	ctctttaaat	2100
tccagctgcg	cggctgggaa	acagcgccac	tgcgcccca	ggccggctgg	aggctgaaga	2160
gcgagctcgc	gctttcgctc	ccggctgcgc	gccgcggaga	gctgggctcg	gcccgcgggc	2220
tgctaggtgg	cggcggcgcg	gggcggggag	gcgcggccc	gcggaggagg	gaagaaaagag	2280
cgagccgggc	cgggagaggc	gccgcgccc	gtcccgcgcc	cgggtcccgca	cccgcctctca	2340
gcggcccaag	cagtttcttt	ctgggtgaca	agaatgtgcc	tcggttggtt	tttctttttt	2400
ttctccatct	ccttaagacg	atttccatag	taacctgatc	aagtggctca	aaatcgcaaaa	2460
cctgaggatt	tccgcggccc	gccggcaaga	cctcggccag	gtaacgctgc	gatctcctcc	2520
tcttccattg	caaaccgctg	cgtccttg	aaagtctctt	ttgtggaaaa	tgcgccagcc	2580
caagggagcc	cggggtat	gcaacagcgt	gttcatttcc	aggtgcctgt	cacgggtctc	2640
ctccctgctg	cttctccagg	acccatgatg	agattat	taaaaaattgt	ttttggtcgt	2700
ctcccccgcc	ccctccctt	ctttat	ttcctcttcg	ctgcactctt	ctcggctttt	2760
cccctgacac	tactgatggg	ggtgcggggg	gacgtcgggg	atgggggtgg	ccagcgcggt	2820
cctgggagtg	gcgggttcgg	atgggctggc	tgcggtgggc	cactttgggc	atctcgcgct	2880
ggcctgcgcc	ggggtcacgg	ggagggctgt	cagcgccagg	gcggcggaac	ccgaggtctc	2940
cagacgagtg	agggagggat	gcaggcttgg	gggtgatgga	gcgcttggct	ggtggctggg	3000
gagcgtccat	acatcatagc	tctccttccc	actcccccg	ccctcttcgg	gattctctct	3060
ttctctttcc	ccgtcctcat	ttctttcttc	ctttactcac	cactcgcttc	attctcttcc	3120
ttccatttcc	tctttttttc	tccctcatt	tccttttttt	cctttccctt	ttaaagaaaag	3180
gggaatcggt	tgtaaccctt	tcgttctacc	aacgtggaat	agctgtgaaa	cctgcagcgt	3240

ggtcacctca gcctggtcgt ttccagaccc gtcccatcc atcaacatat ttgtttcccg 3300  
 agtctattga tctccctgaa ttctacagaa atgcattcta agctaggcgc ctgtatgtca 3360  
 gaatcagttc tgcaggtagc ttccgtgctc caagtatgac atgtattgta agggctgcat 3420  
 ctgttttaaa cccacataag ccatgggtat aaataaatgt agctttgaaa aaaaatctgg 3480  
 ccttattcta gataaacttc cctcttaaat tactgatata ctcttctccc tctttgacat 3540  
 ttaatttttag gaaagttggg agacagggtc ttgtcctcca gtttttaagg agcaggcaac 3600  
 ttctattatc ttaattttct cgtctttgaa catcactcac gtttgcaacta cccagtcagt 3660  
 ggaacgagtg ggtcataatt aa 3682

<210> 14  
 <211> 3904  
 <212> DNA  
 <213> Homo sapiens

<400> 14

cctgcattat tgtttttctc tgacttccaa ttttgggtgt ccctgggtgg gtgggttttc 60  
 ctgacacatt tacaagatgc ttttggcagg ttggctggaa tttgaaggca catttaattg 120  
 taggtgcaat aaaatattca ttttctcttg ttcttgggtt gagatgtcat gcccttttgg 180  
 tcaacttatat tttgggtgtga ctgtgtgtgt gtgtgtatgt gtttgtgtga aggatttaac 240  
 aaagtctgtt ctaactgtca tgtgatttga agttaaaagg tatgttagtg acaagccaca 300  
 aatttctctt atttatagta cattgatect gaaaccattt tttcccttgt gatttcttct 360  
 gtgcatggat catttaacga aagggtggca atgatgagct atttttttat aataggaaaa 420  
 aaattcctca agtttactta ccaagtcata tttttataca gagggattag caaatatttc 480  
 tgatctaata ttttaataga ctgaattgct gaccactgct aattaccaag aatatatttt 540  
 ctttaattctg aaattgctgt acctctcaag ttgtctggag gactccaagt gacccaactt 600  
 gtaactcatg gcaacaggaa gtgggtgttc tgggtgcaag ctgaagtgtg cacatggacc 660  
 cgtactttgt tagcactcgg ggacttgata tggaaagaat taatgtactg gcttttttgt 720  
 atagatgaat gttaactttc tgacattagt cagaactaca tctcccaagc cttgttttgc 780  
 agtgtctgtc cctttgctct tcaactacag taagtcctta ctttaactgac ttgatagggt 840  
 cttggaaact gcaactttta gcaaaaggaa gtataatgaa acacttttat cacaggctaa 900  
 ttggtagaaa caagacttaa gttcccatgg catatttctg gtcacaaaaa catttccaaa 960

cttctcaaaa cacttcaata ttaagcattc aaatacatgt aaactatgta tatatgtaag	1020
aaagggttact ataaaccaga tcaatattta cccaattatt taagttcagg gtcttaggtg	1080
gctggagcct atccgagtag ctcagggcac aaggcgggaa ccagccctag acaggacacc	1140
atcctgttgc agggcacgtt cacacatgcc cacacgcagg ctgggaccat ttacatgtgc	1200
caattcacct accatgcaca tctttgagac gtggcaggaa gcaagagtac ctggagaaaa	1260
tccatacaga tatggggaga atgtacaaac tccaccaga cagtggaccc agccaggaat	1320
caacatttgg gcaacattat aatgaaacga agttgaatga aatgatgtcg ttccacgacc	1380
tgctgtactt gaggggtgtt ataaaattct cagaagacag aggtttaatg ctatcttttt	1440
aatagaaaaat aacttataga gaagtgtgca catgtgactt tgtgtgtagc aggaatcatt	1500
aggatgagaa tcagacgtaa gaggtggtgc caacatgagg aatggtgaga ttcagggagc	1560
tgtggatgga agtagaagcc agaaggccag ggtaggttc ctacttctta ctgtttcagt	1620
tattgcagtg ttggcctgtt tattcacaga tgtcacctag ctttgttttc tcaagaagaa	1680
aatgagcat aatctttcct gttatgaatt cttaaacaca caggacataa ccacagacac	1740
agaggtgcac atatgtagca gtaatggata ctaaatgata cactcggagg aaacagaaaa	1800
gacttctgaa tagagactgg agatacttcc ttggaccatt gatgaatggg caatgatgca	1860
tttttgtctt ccattcagaa ggctaataata ttgctctcta tgttctatgg ataaaggcag	1920
tatatgtctca aggatgaatc acataatatg cataataaat ccagcaagca ttaccctttt	1980
acttatgtga ctgcaagtag gaatacattt cccccactct aaccatgtaa gatttctttc	2040
ccttctccca ttttgtaagc aaaagtaagt tctgaaagg ttaaatggac ctcaggatgg	2100
gaaaaatccc cagagctatc tttctgcaca gacttcattt tttctcccaa gtctgactgt	2160
caactgcgat atctgatatg aggctctggg gctgatgttt ccataggtea tcatccttcg	2220
gtgtcccaga tgaagtctca ggtcgaacat tgcaatagca cagattctga atttaatgca	2280
tcattaaagt tggttatgta acccaatggc cttgttaaac tccagatttt taaaattata	2340
tgtatttact attctcttat tttagaatga tctcacaatg ttcacaagaa ataagcccag	2400
tccctgcaaa gactttaaaa gctgcttggt cacatcatta gattgtacaa cgcttgta	2460
atgacacttt ttgctaactc atgcaacatt tttgtaacaa ttgtgcacat tttactact	2520
tcagataatc aggacctaga gacttcaaga tctggaagca ttgctggtga catagagcaa	2580
aaactttctt gagaatagga agtcagtgtt ttgacaagtg atttataaca gttcaggat	2640

agccaggaag	gtttgaaaca	aaccttaagt	attatttctt	tcactcttgat	tagtatatat	2700
ttatatgtga	tctattttatg	tatatatta	gattttttggg	tcttatagcc	agcttttcatt	2760
tttctctatt	ggaaaagatc	taagtcceca	tccttccttg	gtggcctttg	gtaggtttgt	2820
agacaaaaca	ttgaagaatc	aatggtacct	tttatacatt	aatactgcc	atatgaccat	2880
aaaatcatat	tttttgggaa	tttattcccc	cgatcaaaaag	aagcattttgt	tattgaacac	2940
agtcttatgc	taccttatta	agatgtatca	aacacctga	ttgatcaaaa	acacctcagt	3000
ccattttaag	gcagtattgc	ccagcaatta	aagatgtagc	ttctggagga	gtctttctga	3060
gtttgaattc	agtactcttc	cacgtactat	ataggtgatc	ttgggtaaac	ttcttgagtc	3120
tcagtatccc	catctgtaaa	attgttgtag	agaagaattt	ttgtgatgat	taggtgagag	3180
aatatattaa	tgtaatatatt	aggagagcaa	ccagcatgta	gcataatattc	attacatatc	3240
aatttctata	ttattgatgt	tcatactgct	gatgttgaaa	tgcacaggaa	ggccacagtt	3300
attttctgtt	tagattgatt	tttcttttaa	agtctgaaca	taaactgtaa	tactgtgctt	3360
atztatgtag	gaactgtgat	ctcgtctcct	ccttttcccc	tctccccctc	tctaccttag	3420
tttttcctta	tagtctcaag	ctgaaaacaa	tgaccagggtg	cctaagagat	aagaatactc	3480
tttcttttga	actcatggca	ttagcagtga	cctggatgag	attggagggt	attatttctaa	3540
gtgaaaatagc	tcaggaatgg	aaaaccaagc	attgtatgtt	cttacttata	agtgggagct	3600
aagctatgag	gatacaaagg	cataagaatg	acacaacaga	ctttggagac	ttggggaaaag	3660
ggtgggaagg	gggtgagggga	taaaagacta	caaatagggt	gcagtgtata	ctgcttgggt	3720
ggtgggtgca	ccaaaatctc	acaaatcacc	accaaagaac	ttactcatgt	aaccaaacac	3780
cacctgttcc	ccagtaacct	atggatataa	aaaaattaaa	aaaaagaaaa	aaagaaaact	3840
cttttttgca	gggggcaggt	aaagggtaag	agggcatccc	atttttgagt	ttctagaaaa	3900
gctt						3904

```
<210> 15
<211> 2119
<212> DNA
<213> Homo sapiens
```

<400> 15

ctgcaggaag cagcagcaag gtccagggag cctctaattt aaataggaga agtcagagct 60  
ttaacagcat tgacaaaaac aagcctccaa attatgcaaa tggaaacgaa aaaggttaagt 120

gtttgttaca tcattatgac acaagtccaa catgagtctt gtgaattgca tgctaaatct	180
aatatttgag cagcgtaaca actttgggccc tagagatggt atcagtgagg tttctttatg	240
tttccctaact gtccctcct gactgccagc tttcttatct gaagaacatt ttaaacaagt	300
aaactcatte attttaaagt agttagtatt atatgcaagt acaataactg tttctcaaaa	360
acaggtcctt ccaaatgcat gtaaatcaca ttttcttatg tctttttatg tttttgaaaa	420
tgtatcctga aatcataaag ccatattgaa tttatctgaa tctttaactt cagttaaggt	480
aagagccata agtggttttg acaattaagg ttggagcacc aaaatttgaa acataattac	540
agtaggtttt tatctttgca agcagcagat cccagagata ttatgacctc agttttcccc	600
aaaagacaaa ttattcatat ttgttttggt tctttgaatt agtgcataat ataaatatca	660
aatcacaaaa tcaaggacat taaatgaaag tgtctgttaa aggcataatta taaatgaatc	720
ataagccaca cagttctctg tgatgtacga agtgggcatt taaagagggt ctgatttgat	780
gcttgtcact gagtagcaga gaggacgggg atgagtatgt gtagtttaca cctcaatcat	840
gaggaagtga agaacttggt ctgttataag tagtatggct gtgtgaggaa ctagggtggt	900
ctgctggatt ttgaggaagt attttcaaat caatagaact tcaaaccttt cttcagagtg	960
ttgggctcta catggaaaaa cacatgaaat taaaaagtgg cacaaatggt tagttagtag	1020
aacatctggc taattgggat caaataattc aaccatgtgg gaacgttttt gctcaaaata	1080
gataattgtg aattgtttca tataggcaaa tgattagaca acttcctctt cctcaaatgt	1140
gaacggacag atgtgatcta gaagcaagac actcttttgt gtaaataatt cctttggcct	1200
aaagcaaaag tggacagact ttaaaccact gagagcagag cagtgtgtgt taagattgca	1260
atatcttaag ctcttgagtt aaatggaaaa tgaaaaacaa aagtgtatat ttggaagtta	1320
ggaatgtttt ctttaaaata taaaataaaa ttttagattt aagatcaca gaaatattac	1380
tgaagactta tactcttcct ggggctaagg gaggtgacag tcgctcatca gaaaaaaaaa	1440
aatgccctca tttccctaact tttctaaaaa atataatata agttcaggct aatacttcct	1500
gtatatgtgg gaaatttcta ggggaagcta acaggccttag aaataaagat gtgttaaata	1560
gactaccaa gtgtccaatt aagcaacacg ataccaccgt tattgatatt ctagcaagaa	1620
attactagca atgtttgtaa atagacttag aaatgcattt gatgaattaa cacttttata	1680
tcttaattta tctgaatttt tctgtaatgt gaaaatgttt tatttaactt atttctggca	1740
tctattagta aaattctgat gatatacaag cattaatatt tttccatggc cactcaattc	1800
atacatacct tccctatcta tgcttagaag gcagtgcaaa attagatagt agcaatattg	1860

attataacca caaggtggag acagatgtca tgtaatatgc agtctgctca tataaagcac	1920
atcttcttag acaagagttt tcatacgata taataaagac atctggaatt tgtcttgat	1980
gcaatatgaa atttgctatt aaacgtggag ttaaaacttt atgtcaatag atccaataac	2040
aatgttcata aattaatcat tatgtcatgc tgtatttcca aaatactatc ttaaattata	2100
agagcaaacg aggtaataa	2119

<210> 16  
 <211> 2103  
 <212> DNA  
 <213> Homo sapiens

<400> 16	
gtacatcttt taataaagat gtttgtttta actttttgaa tatgaagatt tctagttcta	60
gaataatggt tataaaaata tacaatcca tctggtgatg agttgacctc taccacaact	120
agtttgcata tataacttgg gtgtgaccaa gcaaggtgag agttaagaac ttttaaaact	180
tactgtatta tattgataga actcagaaag tactaaacttg aatattatta ttctaattgc	240
ttttcccttt tagttattaa aaataagaat acttaaatta ataacaagat cttttactgg	300
caggattaac caaattatct gtaatgtgtt cctcgaatgc ttttaagtgg aaatatactt	360
tatacattct ttaacaactc tgagaggatg agttacataa atcagttcag gaatctatag	420
aatctgtaat acatagtaaa ggtttattca caattaaaac aatttcactt ctatattaaa	480
aaaacaaatt gttgaaagta cagtggcttt tcatatgtat gatttgtaaa acaaattagc	540
ttttttaaag tgatgtgacg cttaatgaga agaaatcagt agagaattac aaactgcact	600
tcaaaagata catctaatat cattttaata atgaaatttg aaaaaatagt gtgctcgttt	660
tacagtctca ttaaatgaat taaaatatca gcacacattg tagtaggtta tcattggcag	720
agaaggctga aatagaaacg ttacaatggg atgcactgcc atctgaacat tatgtcgaag	780
tggaacgcgg aaacatattt ctcagaacaa gtggtaaaat gaaaacagca tcatttgtaa	840
agcatttctt ttgagagtgc ttcagtttct tctcctgatg acctgccatt cagaaactga	900
caatgaataa tacactctga caccagcatt tgtcaatttg cccagaacca tatgagagta	960
ctctagacag atatatgttc cgaagtaaac cgaatacctg ttaactgtaa atcaaatctt	1020
gtagaaacca tgccatgggt cctttggaca tatactttgc atgctgaag caagttacct	1080
taagaaatca ttcttttggt ttacaaaact tgtattaaaa aattaaaaat gcaaaaaagc	1140

ttaatatattat	taggaatttta	tccatagctt	tattttggaat	ccagttttctt	tattatgatc	1200
tataaacatg	catcatttga	tggagttcct	tagtggagag	gtgtttttcc	atgttgctaa	1260
gaaacatgcc	ccagcaccag	aagggatact	acctaccatc	tttttgccat	ttctcaccgt	1320
gattcttaca	ttgtacctgt	ttactcactg	aacagggctt	ccttctcttt	gtctagattc	1380
taatcaggtg	tcttctggtg	tggaagcttt	ggctttttatt	tacacacaac	acagaattaa	1440
taagatagat	gccaaggatt	tagcaacatt	ttaattcaac	attatacagg	tatcagagtt	1500
aatgagaatt	atgcattagt	ctttaaattt	gggcagctta	ttcagctaaa	acatagatgt	1560
ctagctctta	aacactttgt	ttttttaatt	actctgaaat	tacaataaaag	tcaaagaact	1620
gaactgtttt	cttttcaagc	cagtgc aaat	gtgctttagt	tattatttta	ctggtgatct	1680
aattatgcat	tttaatgctt	tattacttaa	tacttatata	agcctaaaaat	acgttgttaa	1740
tgtcataaatt	tcaggggattt	tagtattctt	tccatgagtt	accataacta	ggtgcatatg	1800
tgtaaataata	cgtatatatc	tatatctata	tatttatatc	tatgtatatata	tcaatttata	1860
agactaaata	gacttggcca	tatgtgttgt	tggtttatgc	atacatgcac	aaatattgag	1920
gtgtccacaa	agtatatatg	cctgtacata	aattacatac	tggctgggtga	gtgaatgtaa	1980
gcttctctaa	attgtacaac	tctccacaga	gtggcactct	aatattgcaa	aggtacaata	2040
taagcatgtg	cagaatgaac	agctcttcta	ggatccctat	aaaactccac	cccatgtttc	2100
tgt						2103

```
<210> 17
<211> 4260
<212> DNA
<213> Homo sapiens

<400> 17
```

aagcttcac	ccagagggg	acttgccaga	tgcttgctag	agctctcctg	tatgaggagt	60
ctatcaacac	ctgctgggag	gtgtctcctc	gtcaggaggc	acgggggtca	gggaccact	120
tgaggaggct	gtctgtccct	tagcggagct	agaacactgt	gctcggagat	ccgctgctct	180
cttcagagct	ggcaggcaag	agtgttttag	tctgtctgagc	ctgcgcccac	agccgcccct	240
tccccaggt	gctctgtccc	agggagatga	gagttttatc	tgtaagcccc	tgactggggc	300
tgctaccttt	ctttcagata	tgccccgccc	agagaggagg	aatctagaga	ggcagtctgg	360
ctacagcagc	tttgccaagc	tgcagtgggc	tctgcccagt	ccaaaattcc	cagcgggttt	420

gtttacattg tgaggggaaa agcacctact caagcctcag ttatggcagt tgccccctccc	480
cccaccaagc tccaggggtcc caggtgtcct tcagactgct gtgctggcaa tgagaatttc	540
aagccagtgg atcttagctt gctgggctcc acaggggtgg gatccactga gctagaccac	600
ttagctccct ggcttcagcc ccccttcag gtgagtggat ggctctgtct cactggcatt	660
ccaggtgcta ctgggggtatg aaaaaaaaaa ctctgcagc tagcttggtg tctgccagt	720
tttgtgcttg aaactcaggc ccttggtggt gtggacaccc aatggaatct cctggtgtgc	780
atgttgatga gactgtggga aaagcatagt atctgggctg gatagctccg tccttcaagg	840
cacagtccct catgacttcc cttggctagg ggagggagtt cccaaccct ttgcacttcc	900
caggtgaggc aacaccccac cctgcttctg ctccacctct gtgggctgca cccactgtct	960
aatcagtcac tgtgagatga gcctgggtacc tcagttggaa atgcagaaat cacctgcctt	1020
ctgtgttgat ctactggga gcagcagact ggagctgttc ctattcagcc atctttctca	1080
ggtcataatc atagattttt aattgatccc agcaacatgg attagtaaac agcatatttc	1140
caagtgattt ttttttattt taaggtcaaa tctacaaaat attatagtgt taccaccact	1200
taaaattatt actggtgata ctatgtttgt ctctattcac attttattgc tagaaagaat	1260
tataatttgt agataataat agttatttga aatgtattac atatcctttt acttttaaga	1320
agaggtgact taattatcta ggtatacaat tttttgagg atactaaatg tcatgaatag	1380
caaatttatc atattgcttt cctaggtgaa gacctgaaa caagaagaat gagaacagtt	1440
aaaaacatag cagacttgag gcagaattta gaagagacta tgtccagtct tcgtgggact	1500
cagataagcc acaggttttt ttcaattttg catatatattg agccaataaa gaaaaaataa	1560
ttacaaacaa acatttaact tttcttataa tgacagagat gggatttcag tttcccccta	1620
ctattttctc ccttgtttta tatcaaattg attggtaatt atccttaaag tgagaattca	1680
cagtatatac ctatttatct tttatctcta tctctatctg ctatttatgt ctttttcagt	1740
ataatttcca gtactgcaac taccaccatc actgttaagt ggatttgtaa tacctgtcct	1800
agaaaacagt ggcacaagtt gcacttgaaa tgcactctgg cagggtagta gggagacatt	1860
caaacataat tgtagttaac tttcagaata ggtctgggaa ggttacagtg agttaaggat	1920
ttgttgaaaa tgtaaaacaa tatgttggtt tacccaaggt gtactgatgg cttttctttt	1980
gaaaacaaac gaaaagctat aaaatgtatg cccctttcca caatttgacc tcaaaatgaa	2040
tatagagttt agctttcggg aagatgacgt gtttataaga gatgaccctc aactccagcc	2100



ttttctgtct	tcattgcattc	tagatttatgg	ccctaagtga	accagagtat	agttattttct	2160
ccatttttatt	tgacagcacc	ctggagacaa	catttgacag	cactgtgaca	acagaagtta	2220
atggaaggac	catacccaac	ttgacaagtc	gacccacccc	catgacctgg	aggttggggcc	2280
aggcatgtcc	gcgacttcag	gcgggagatg	ctccctccct	gggtgctggc	tatcctcgca	2340
gtggtaccag	tcgattcatc	cacacagacc	cctcgagggt	catgtatacc	acgcctctcc	2400
gtcgagctgc	tgtctctagg	ctgggaaaca	tgtcacagat	tgacatgagt	gagaaagcaa	2460
gcagtgacct	ggacatgtct	tctgaggtcg	atgtgggtgg	atatatgagt	gatggtgata	2520
tccttgggaa	aagtctcagg	actgatgaca	tcaacagtgg	gtaagtaacc	ctgttctccg	2580
tcagcattgt	gtgaagaggg	gaggtggtct	actataatgc	attcactata	aacaaatgtg	2640
taagtttgcc	cagaaagtca	tgagaacata	tgagatatct	gaggttattc	agagtgttga	2700
agggcccttc	ctctgctcat	tcattggagag	taaagaatcc	aagattttcta	taaattcatt	2760
ataagccgct	aagtttttct	gttgttgaga	gaaacacatg	tggcttctgt	ttttcagagt	2820
gattttcaca	tgcttcttaa	gtaacagatt	ttgtagttaa	ggacgtggga	aggagacagg	2880
aggagttttg	ctgatttgct	tgattttttt	tttctttttt	agcttggttag	aagcggcctg	2940
taactgcttt	gagaaacaaa	tattttctta	ctgtcttcaa	ttatgcatcc	ccaatttaac	3000
ttgagggaaa	aatcactttg	gagttgaaag	tttactcta	ttcattttct	tttgatggta	3060
tcagatttca	atacatctca	gacctgtttt	ttcttctgtg	tcctattaca	ttccaaaaca	3120
tgttgtgatt	gtaaaactct	tagagtatat	taacaatttg	ggatatttgg	cataatcaga	3180
gaataggtcc	aaaaggaggc	aataggatat	tctattaata	attgtaattg	ccatttttag	3240
catttcctgt	tatgtactat	gctcttgtca	agtgccttga	agatagtgtt	ttacttttcc	3300
ttcccaccac	cagcaatgtt	tatgaggtag	atgtttttat	acatgttcta	tggataagga	3360
aactgagtct	aattggcccc	ggctgggaac	taacgctagg	gaaacggcag	acctgcatta	3420
gaactcagct	atgtctgact	tcaaacacag	gctcagtaat	atgtggaaaa	gcttcccaat	3480
taactttgtc	tataaacttt	gtgtgagtct	ggattttgac	ttactctttg	tctttacgca	3540
tctgagagga	cccatgtagg	aaataattct	tctatataag	tgacccttcc	tgacttcatt	3600
catgaaaagc	ttatgtttga	agggtgacac	gacctaaaaa	agagtacaaa	atagcttttg	3660
attacattta	tagctttgct	ctgatatcct	aatacctact	agtccattcc	tggtatccac	3720
cctacctgac	tttctaaaaa	tttagaatta	tagagactaa	ttatgattaa	ttaagatagg	3780
ttgttgttca	gttgccactg	gattcagagt	gcctagtttg	aatctctccc	attcactatc	3840

tgtggacccc ttcggaacct aacgtatcca aattagtttt tgatcatctag aataaggata 3900  
 aaattgtacc atcttcatga agttgttagg atcatccaca aatttttagtt tgcgcaatgc 3960  
 ttggcatgat acaagcactc aataaattta tcatcttcct ctttatcatc actattacat 4020  
 ttattatcat taataaccat accaattttt gggtgtgtgt agttataatt atcatttttg 4080  
 tatgtattta acatagccta ggaggcaatg cccagttcag aaaacataat ggcaaagcaa 4140  
 gagtgtctaa ggcacactct ttctcccatc tctctcttct ttcttctcca ttctttccac 4200  
 tctatcccct cttctctttt ttttctcaat ctcttagat gtggacatat gtgtgaattc 4260

<210> 18

<211> 2429

<212> DNA

<213> Homo sapiens

<400> 18

tgtgggtgtg ggtgtgaagc atgtgtatgt gtgtgtgtga agcatctccc cacctgtaat 60  
 gtaagtccat gagtgcagaa tttttgacat attctttacg tgttgagttt taacaaatgt 120  
 ttgtggagtg aatgaacaaa ttaatgaata taggctattht attaattagg caatatagtc 180  
 acataggctg gcaatcgcac ctaattaaat agagtggtaa atgagttcca gaaagaacta 240  
 aggtactaca aggatgttat gaaagagaaa aatgagttat gtgaaaaata ggagacagtg 300  
 ataagagggg aagaatccca aagtgtgggc cacatthtga aactaatgac ctattattct 360  
 attattgtta gctgaaagta gaaaacgtca tgggagggaa tatctgctag tttttggtaa 420  
 aggatgttgt gatggcagaa ccaagaaatg aacacaaggt gactttgggt tggggacagt 480  
 gggataatca actctccttg ctccatcagg gccccagact gggctctggc agaggaactc 540  
 agaacaacgt aaagacctag ataggtatct aataaattgg gacctgtgaa aacagtgcct 600  
 cttaaagtgt ggtacctgga ccagcagcag cagcagcagc agccattgaa acttcataga 660  
 aagacagatt ctacgttca tccaagactt actgaattag aatatctcaa ggtaaggcct 720  
 ggtaatctga gctttaacta gccctcaagg tgattcttaa gttcaagcat cactatatta 780  
 agttgaacaa atagatgcca ggcctataaa tacatgtaac gcctagcata aatatttcaa 840  
 cattaaaaat gacatttcat agttcttatt taccctatta gctgtgttct gtcaagataa 900  
 tgagaatatt gatatgtag aatacactga tgcactaatt tttaaattag atcaaataat 960  
 gacttggtat acctgaaata aattgggtca gcttggtaga tgcagthttht gagaattata 1020

taagtcattt	ttaaaagaat	aatttttaact	tgagctgctt	gcataaatta	aattgcaaaa	1080
aggtcatagt	ataaatcctc	ctatttagcag	agatagaagg	tttttaaaaa	aattacagat	1140
aagtctgaag	gtctttttaa	atctttatatt	caggaagtga	ctcgggatgt	atatcatttt	1200
aaaatacatg	gtctttaaag	ttgtagttgt	atgactcttt	cagttaattt	aaaatacttc	1260
cttctatgaa	aaattgtttc	aaaaattttt	ctaaattctg	ttatccattt	caagtaggat	1320
aggcaagaac	agatataaga	tactactttt	ttgttcatgt	ttactaaaaa	aaaaattact	1380
gtaattgaga	tcattgtaaaa	acattgtttc	tgtctatttg	tcttaacctt	ttaatcctgg	1440
caccttaaat	ttgacatagt	aggaattaga	agacaattgc	agaaaatgtc	aactggggaa	1500
attttattct	actaaaaact	atgtccatac	aacatagcaa	atcacatttt	aaaggccaaa	1560
aagtctttca	tagcaatttt	tcagattatt	ttcaaagcat	atcttctctc	tgtctctgca	1620
gcatgccgtt	gatttttctg	ttatgcagtc	acataagtaa	ttacatgttt	acattgtctat	1680
ttcactcata	gaacacgaaa	cagttaaagt	tagaataata	tccaatccat	ctttttatca	1740
ccagtagcta	gcatactgta	ggaactcaat	aaatatatca	gataaattgt	ggaaataacc	1800
atatcagctt	ataacatata	gaaatgtgag	tttaaaaaga	aaacaattat	acatatgaaa	1860
aaatttttat	accatttttt	taaagacctt	tcagatgtca	tacagtttgg	acttttccag	1920
tgtttcttgt	atcatgagac	aatagtagac	attgtaaatc	aaaaatagtt	ttctgggggt	1980
gtgtacattt	gaaaaaactg	aatatcatat	ctgttcttag	agagtaatga	tggatattaa	2040
catatcaaag	gtacagagaa	gtcttaaagt	tcaaagtaac	atctgcttaa	ttgtatttaa	2100
ttcagtgtct	catgagcttt	tttatcactg	attccctccc	tttttctctt	tatgataata	2160
attaacttgt	tcctgtagca	ttttaagaaa	tgttgattta	gttgaatgcc	ttcacttctc	2220
caatataata	gcagaaactc	agaaatatatt	atttaccag	aatcatgcag	ctaatagtac	2280
aaggattcag	gtcttttact	tcctattttg	tggttcccaa	ctacttttgc	caaaggctct	2340
ttaaataata	tgaacatat	tagtgattga	ttcattatag	taaatgggta	aatgataagg	2400
cttgcaataa	ttcactgaca	agaaagctt				2429

<210>	19
<211>	2926
<212>	DNA
<213>	mouse

<220>  
<221> CDS

&lt;222&gt; (2) ... (2926)

&lt;400&gt; 19

a agc cac agc acc ctg gag aca acc ttt gat acg act gtg aca act Ser His Ser Thr Leu Glu Thr Thr Phe Asp Thr Thr Val Thr Thr	46
gaa gtg aat gga agg gcc atc ccc aac ctg aca agc cga cct tcc ccc Glu Val Asn Gly Arg Ala Ile Pro Asn Leu Thr Ser Arg Pro Ser Pro	94
atg acc tgg aga ctg ggt caa gcg tgc cct cgt cta cag gct gga gat Met Thr Trp Arg Leu Gly Gln Ala Cys Pro Arg Leu Gln Ala Gly Asp	142
gcc ccc tcc atg ggc gct gga tat tct cga agc ggt acc agc cga ttc Ala Pro Ser Met Gly Ala Gly Tyr Ser Arg Ser Gly Thr Ser Arg Phe	190
atc cac acg gat ccc tcc agg ttt atg tat acc acg cct ctc cgc cga Ile His Thr Asp Pro Ser Arg Phe Met Tyr Thr Thr Pro Leu Arg Arg	238
gct gct gtc tcg cgt ctg gga aac atg tca caa ata gat atg agc gag Ala Ala Val Ser Arg Leu Gly Asn Met Ser Gln Ile Asp Met Ser Glu	286
aaa gca agc agt gac ctg gat gtg tct tct gaa gtg gat gtt ggt gga Lys Ala Ser Ser Asp Leu Asp Val Ser Ser Glu Val Asp Val Gly Gly	334
tac atg agc gat ggt gat atc ctt ggg aag agt ctg aga gcg gat gat Tyr Met Ser Asp Gly Asp Ile Leu Gly Lys Ser Leu Arg Ala Asp Asp	382
atc aac agt ggg tac atg aca gat ggt ggg ctc aac cta tat acc aga Ile Asn Ser Gly Tyr Met Thr Asp Gly Gly Leu Asn Leu Tyr Thr Arg	430
agt ctt aac cga gtc ccg gac aca gca act tcc aga gat gtc ata cag Ser Leu Asn Arg Val Pro Asp Thr Ala Thr Ser Arg Asp Val Ile Gln	478
aga ggc gtt cac gat gtg aca gtg gac gca gac agc tgg gat gac agc Arg Gly Val His Asp Val Thr Val Asp Ala Asp Ser Trp Asp Asp Ser	526
agt tct gtg agc agt ggc ctc agt gac aca ctt gat aac att agc aca Ser Ser Val Ser Ser Gly Leu Ser Asp Thr Leu Asp Asn Ile Ser Thr	574
gat gac ctc aac acc acg tcc tcc atc agt tct tac tcc aac atc act Asp Asp Leu Asn Thr Thr Ser Ser Ile Ser Ser Tyr Ser Asn Ile Thr	622

gtc ccc tcc agg aag aac act cag ctg aaa aca gat gcg gag aaa cgt Val Pro Ser Arg Lys Asn Thr Gln Leu Lys Thr Asp Ala Glu Lys Arg	670
tcg aca aca gat gag acc tgg gat agt cct gag gag ctg aag aaa gcc Ser Thr Thr Asp Glu Thr Trp Asp Ser Pro Glu Glu Leu Lys Lys Ala	718
gag gga gat tgt gac agc cat ggt gac gga gcc gcc aag tgg aag ggt Glu Gly Asp Cys Asp Ser His Gly Asp Gly Ala Ala Lys Trp Lys Gly	766
gct act tct gga ctt gct gaa gac tcg gag aag aca ggg cag aaa gcc Ala Thr Ser Gly Leu Ala Glu Asp Ser Glu Lys Thr Gly Gln Lys Ala	814
agc ctg tct gtg tct cag aca ggc tcc tgg agg aga ggc atg tct gcc Ser Leu Ser Val Ser Gln Thr Gly Ser Trp Arg Arg Gly Met Ser Ala	862
cag gga gga act cca gct aca gct agg cag aaa acc agc aca agt gca Gln Gly Gly Thr Pro Ala Thr Ala Arg Gln Lys Thr Ser Thr Ser Ala	910
ctc aag acc cct ggg aag aca gat gat gcc aaa gct tcc gag aaa ggg Leu Lys Thr Pro Gly Lys Thr Asp Asp Ala Lys Ala Ser Glu Lys Gly	958
aaa act cct ctc aaa gga tca tcc ttg caa agg tct cct tca gat gca Lys Thr Pro Leu Lys Gly Ser Ser Leu Gln Arg Ser Pro Ser Asp Ala	1006
ggg aaa agc agc ggg gat gaa ggg aaa aag cca ccg tca ggc att gga Gly Lys Ser Ser Gly Asp Glu Gly Lys Lys Pro Pro Ser Gly Ile Gly	1054
aga tcg aca gcc agc agt tct ttt gga tac aag aag cca agt ggt gta Arg Ser Thr Ala Ser Ser Ser Phe Gly Tyr Lys Lys Pro Ser Gly Val	1102
ggg gct tcc act atg att acc agc agc ggt gcc acc atc aca agc ggt Gly Ala Ser Thr Met Ile Thr Ser Ser Gly Ala Thr Ile Thr Ser Gly	1150
tca gct aca ctg ggg aaa atc ccc aaa tcc gct gcc att ggt ggg aag Ser Ala Thr Leu Gly Lys Ile Pro Lys Ser Ala Ala Ile Gly Gly Lys	1198
tcc aat gca gga agg aaa acc agc ctg gac ggg tcc cag aat caa gat Ser Asn Ala Gly Arg Lys Thr Ser Leu Asp Gly Ser Gln Asn Gln Asp	1246
gat gtt gtc ctg cac gtg agc tcg aag acc acc ctc cag tac cgt agt Asp Val Val Leu His Val Ser Ser Lys Thr Thr Leu Gln Tyr Arg Ser	1294
ttg ccc cgc cct tct aag tcc agc acc agc gga atc cct ggg aga ggt Leu Pro Arg Pro Ser Lys Ser Ser Thr Ser Gly Ile Pro Gly Arg Gly	1342

ggc cac agg tcg agc acc agc agc att gat tcc aat gtc agc agc aag Gly His Arg Ser Ser Thr Ser Ser Ile Asp Ser Asn Val Ser Ser Lys	1390
tca gct ggg gcc acc acc tcc aaa ctg aga gaa ccg act aag atc ggc Ser Ala Gly Ala Thr Thr Ser Lys Leu Arg Glu Pro Thr Lys Ile Gly	1438
tca ggg cgc tcg agt cca gtc act gtc aac caa aca gac aaa gag aag Ser Gly Arg Ser Ser Pro Val Thr Val Asn Gln Thr Asp Lys Glu Lys	1486
gag aaa gta gca gtg tca gat tca gag agc gtt tcc ttg tca ggt tcc Glu Lys Val Ala Val Ser Asp Ser Glu Ser Val Ser Leu Ser Gly Ser	1534
ccc aaa tcc agc ccc acc tct gcc agt gcc tgt ggg act caa ggg ctc Pro Lys Ser Ser Pro Thr Ser Ala Ser Ala Cys Gly Thr Gln Gly Leu	1582
aga cag cca ggg tcc aaa tat cca gat att gcc tcg ccc aca ttt cga Arg Gln Pro Gly Ser Lys Tyr Pro Asp Ile Ala Ser Pro Thr Phe Arg	1630
agg ttg ttc ggt gcc aag gca ggc ggc aaa tct gcc tcc gca cct aat Arg Leu Phe Gly Ala Lys Ala Gly Gly Lys Ser Ala Ser Ala Pro Asn	1678
act gag ggg gcg aag tcc tcc tca gta gtg ctc agc cct agt acc tct Thr Glu Gly Ala Lys Ser Ser Ser Val Val Leu Ser Pro Ser Thr Ser	1726
tta gcc cga caa ggc agt ctg gag tca ccg tcg tcc ggt acg gga agc Leu Ala Arg Gln Gly Ser Leu Glu Ser Pro Ser Ser Gly Thr Gly Ser	1774
atg ggc agt gct ggt ggg ctg agt ggc agc agc agc cct ctc ttc aat Met Gly Ser Ala Gly Gly Leu Ser Gly Ser Ser Ser Pro Leu Phe Asn	1822
aaa ccc tca gac cta act aca gat gtt ata agc tta agt cac tcc ttg Lys Pro Ser Asp Leu Thr Thr Asp Val Ile Ser Leu Ser His Ser Leu	1870
gct tcc agc cca gcg tcg gtt cac tct ttc aca tcc ggt ggg ctt gtg Ala Ser Ser Pro Ala Ser Val His Ser Phe Thr Ser Gly Gly Leu Val	1918
tgg gct gcc aat ctg agc agt tcc tct gcc ggc agc aag gac act cca Trp Ala Ala Asn Leu Ser Ser Ser Ser Ala Gly Ser Lys Asp Thr Pro	1966
agt tac cag tcc atg act agt ctc cat acg agc tct gag tcc att gac Ser Tyr Gln Ser Met Thr Ser Leu His Thr Ser Ser Glu Ser Ile Asp	2014

ctg	ccc	ctc	agc	cat	cat	ggc	tcc	ctg	tct	gga	ctg	acc	aca	ggc	act	2062
Leu	Pro	Leu	Ser	His	His	Gly	Ser	Leu	Ser	Gly	Leu	Thr	Thr	Gly	Thr	
cac	gag	gtg	cag	agc	ctg	ctc	atg	aga	acg	ggt	agt	gtg	aga	tct	act	2110
His	Glu	Val	Gln	Ser	Leu	Leu	Met	Arg	Thr	Gly	Ser	Val	Arg	Ser	Thr	
ctc	tca	gaa	aga	tac	acc	cca	tca	tct	cgg	cag	gcc	aac	caa	gaa	gaa	2158
Leu	Ser	Glu	Arg	Tyr	Thr	Pro	Ser	Ser	Arg	Gln	Ala	Asn	Gln	Glu	Glu	
ggc	aaa	gag	tgg	ctg	cga	tcg	cat	tcc	act	ggc	ggg	ctg	cag	gat	act	2206
Gly	Lys	Glu	Trp	Leu	Arg	Ser	His	Ser	Thr	Gly	Gly	Leu	Gln	Asp	Thr	
ggc	aac	cag	tct	ccc	ttg	gtc	tcc	cct	tct	gcc	atg	tca	tcg	tca	gcc	2254
Gly	Asn	Gln	Ser	Pro	Leu	Val	Ser	Pro	Ser	Ala	Met	Ser	Ser	Ser	Ala	
acc	gga	aaa	tat	cac	ttt	tcc	aac	ttg	gtg	agt	ccc	acc	aac	ctc	tcc	2302
Thr	Gly	Lys	Tyr	His	Phe	Ser	Asn	Leu	Val	Ser	Pro	Thr	Asn	Leu	Ser	
cag	ttt	aac	ctg	cct	gca	ccc	agt	atg	atg	cgc	tcc	agc	agt	atc	ccc	2350
Gln	Phe	Asn	Leu	Pro	Ala	Pro	Ser	Met	Met	Arg	Ser	Ser	Ser	Ile	Pro	
gcc	cag	gac	tcc	tcc	ttc	gac	ctc	tat	gat	gat	gcc	cag	ctt	tgc	ggc	2398
Ala	Gln	Asp	Ser	Ser	Phe	Asp	Leu	Tyr	Asp	Asp	Ala	Gln	Leu	Cys	Gly	
agt	gca	act	tcc	ctg	gag	gaa	agg	cca	cgg	gcc	gtt	agc	cac	tcc	ggc	2446
Ser	Ala	Thr	Ser	Leu	Glu	Glu	Arg	Pro	Arg	Ala	Val	Ser	His	Ser	Gly	
tca	ttc	aga	gac	agc	atg	gag	gaa	gtt	cat	ggc	tct	tca	ctg	tca	ttg	2494
Ser	Phe	Arg	Asp	Ser	Met	Glu	Glu	Val	His	Gly	Ser	Ser	Leu	Ser	Leu	
gtc	tcc	agc	aca	tca	tcc	ctt	tac	tct	acg	gct	gaa	gag	aag	gct	cat	2542
Val	Ser	Ser	Thr	Ser	Ser	Leu	Tyr	Ser	Thr	Ala	Glu	Glu	Lys	Ala	His	
tca	gag	caa	atc	cat	aag	cta	cgg	aga	gaa	ctg	gtt	gcc	tcc	cag	gag	2590
Ser	Glu	Gln	Ile	His	Lys	Leu	Arg	Arg	Glu	Leu	Val	Ala	Ser	Gln	Glu	
aaa	gtc	gct	acc	ctc	acg	tct	cag	ctg	tca	gca	aat	gct	cac	ctt	gta	2638
Lys	Val	Ala	Thr	Leu	Thr	Ser	Gln	Leu	Ser	Ala	Asn	Ala	His	Leu	Val	
gca	gct	ttt	gaa	aag	agt	tta	ggg	aat	atg	act	ggc	cgt	ttg	caa	agt	2686
Ala	Ala	Phe	Glu	Lys	Ser	Leu	Gly	Asn	Met	Thr	Gly	Arg	Leu	Gln	Ser	
cta	acc	atg	aca	gcg	gaa	caa	aag	gaa	tct	gag	ctt	atc	gaa	ctg	cgg	2734





Gly	Val	His	Asp	Val	Thr	Val	Asp	Ala	Asp	Ser	Trp	Asp	Ser	Ser	
				165					170					175	
Ser	Val	Ser	Ser	Gly	Leu	Ser	Asp	Thr	Leu	Asp	Asn	Ile	Ser	Thr	Asp
			180					185					190		
Asp	Leu	Asn	Thr	Thr	Ser	Ser	Ile	Ser	Ser	Tyr	Ser	Asn	Ile	Thr	Val
		195					200					205			
Pro	Ser	Arg	Lys	Asn	Thr	Gln	Leu	Lys	Thr	Asp	Ala	Glu	Lys	Arg	Ser
	210					215					220				
Thr	Thr	Asp	Glu	Thr	Trp	Asp	Ser	Pro	Glu	Glu	Leu	Lys	Lys	Ala	Glu
225					230					235					240
Gly	Asp	Cys	Asp	Ser	His	Gly	Asp	Gly	Ala	Ala	Lys	Trp	Lys	Gly	Ala
				245					250					255	
Thr	Ser	Gly	Leu	Ala	Glu	Asp	Ser	Glu	Lys	Thr	Gly	Gln	Lys	Ala	Ser
			260					265					270		
Leu	Ser	Val	Ser	Gln	Thr	Gly	Ser	Trp	Arg	Arg	Gly	Met	Ser	Ala	Gln
		275					280					285			
Gly	Gly	Thr	Pro	Ala	Thr	Ala	Arg	Gln	Lys	Thr	Ser	Thr	Ser	Ala	Leu
	290					295					300				
Lys	Thr	Pro	Gly	Lys	Thr	Asp	Asp	Ala	Lys	Ala	Ser	Glu	Lys	Gly	Lys
305					310					315					320
Thr	Pro	Leu	Lys	Gly	Ser	Ser	Leu	Gln	Arg	Ser	Pro	Ser	Asp	Ala	Gly
				325					330					335	
Lys	Ser	Ser	Gly	Asp	Glu	Gly	Lys	Lys	Pro	Pro	Ser	Gly	Ile	Gly	Arg
			340					345					350		
Ser	Thr	Ala	Ser	Ser	Ser	Phe	Gly	Tyr	Lys	Lys	Pro	Ser	Gly	Val	Gly
		355					360					365			
Ala	Ser	Thr	Met	Ile	Thr	Ser	Ser	Gly	Ala	Thr	Ile	Thr	Ser	Gly	Ser
	370					375					380				
Ala	Thr	Leu	Gly	Lys	Ile	Pro	Lys	Ser	Ala	Ala	Ile	Gly	Gly	Lys	Ser
385					390					395					400
Asn	Ala	Gly	Arg	Lys	Thr	Ser	Leu	Asp	Gly	Ser	Gln	Asn	Gln	Asp	Asp
				405					410					415	
Val	Val	Leu	His	Val	Ser	Ser	Lys	Thr	Thr	Leu	Gln	Tyr	Arg	Ser	Leu
			420					425					430		
Pro	Arg	Pro	Ser	Lys	Ser	Ser	Thr	Ser	Gly	Ile	Pro	Gly	Arg	Gly	Gly
		435					440					445			
His	Arg	Ser	Ser	Thr	Ser	Ser	Ile	Asp	Ser	Asn	Val	Ser	Ser	Lys	Ser

450		455		460
Ala Gly Ala Thr Thr Ser Lys Leu Arg Glu Pro Thr Lys Ile Gly Ser				
465		470		475
Gly Arg Ser Ser Pro Val Thr Val Asn Gln Thr Asp Lys Glu Lys Glu				
	485		490	495
Lys Val Ala Val Ser Asp Ser Glu Ser Val Ser Leu Ser Gly Ser Pro				
	500		505	510
Lys Ser Ser Pro Thr Ser Ala Ser Ala Cys Gly Thr Gln Gly Leu Arg				
	515		520	525
Gln Pro Gly Ser Lys Tyr Pro Asp Ile Ala Ser Pro Thr Phe Arg Arg				
	530		535	540
Leu Phe Gly Ala Lys Ala Gly Gly Lys Ser Ala Ser Ala Pro Asn Thr				
545		550		555
Glu Gly Ala Lys Ser Ser Ser Val Val Leu Ser Pro Ser Thr Ser Leu				
	565		570	575
Ala Arg Gln Gly Ser Leu Glu Ser Pro Ser Ser Gly Thr Gly Ser Met				
	580		585	590
Gly Ser Ala Gly Gly Leu Ser Gly Ser Ser Ser Pro Leu Phe Asn Lys				
	595		600	605
Pro Ser Asp Leu Thr Thr Asp Val Ile Ser Leu Ser His Ser Leu Ala				
	610		615	620
Ser Ser Pro Ala Ser Val His Ser Phe Thr Ser Gly Gly Leu Val Trp				
625		630		635
Ala Ala Asn Leu Ser Ser Ser Ser Ala Gly Ser Lys Asp Thr Pro Ser				
	645		650	655
Tyr Gln Ser Met Thr Ser Leu His Thr Ser Ser Glu Ser Ile Asp Leu				
	660		665	670
Pro Leu Ser His His Gly Ser Leu Ser Gly Leu Thr Thr Gly Thr His				
	675		680	685
Glu Val Gln Ser Leu Leu Met Arg Thr Gly Ser Val Arg Ser Thr Leu				
	690		695	700
Ser Glu Arg Tyr Thr Pro Ser Ser Arg Gln Ala Asn Gln Glu Glu Gly				
705		710		715
Lys Glu Trp Leu Arg Ser His Ser Thr Gly Gly Leu Gln Asp Thr Gly				
	725		730	735
Asn Gln Ser Pro Leu Val Ser Pro Ser Ala Met Ser Ser Ser Ala Thr				
	740		745	750

Gly	Lys	Tyr	His	Phe	Ser	Asn	Leu	Val	Ser	Pro	Thr	Asn	Leu	Ser	Gln
755						760						765			
Phe	Asn	Leu	Pro	Ala	Pro	Ser	Met	Met	Arg	Ser	Ser	Ser	Ile	Pro	Ala
770						775						780			
Gln	Asp	Ser	Ser	Phe	Asp	Leu	Tyr	Asp	Asp	Ala	Gln	Leu	Cys	Gly	Ser
785						790						795			
Ala	Thr	Ser	Leu	Glu	Glu	Arg	Pro	Arg	Ala	Val	Ser	His	Ser	Gly	Ser
				805						810				815	
Phe	Arg	Asp	Ser	Met	Glu	Glu	Val	His	Gly	Ser	Ser	Leu	Ser	Leu	Val
				820				825				830			
Ser	Ser	Thr	Ser	Ser	Leu	Tyr	Ser	Thr	Ala	Glu	Glu	Lys	Ala	His	Ser
		835						840				845			
Glu	Gln	Ile	His	Lys	Leu	Arg	Arg	Glu	Leu	Val	Ala	Ser	Gln	Glu	Lys
850						855						860			
Val	Ala	Thr	Leu	Thr	Ser	Gln	Leu	Ser	Ala	Asn	Ala	His	Leu	Val	Ala
865						870				875				880	
Ala	Phe	Glu	Lys	Ser	Leu	Gly	Asn	Met	Thr	Gly	Arg	Leu	Gln	Ser	Leu
				885						890				895	
Thr	Met	Thr	Ala	Glu	Gln	Lys	Glu	Ser	Glu	Leu	Ile	Glu	Leu	Arg	Glu
		900						905				910			
Thr	Ile	Glu	Met	Leu	Lys	Ala	Gln	Asn	Ser	Ala	Ala	Gln	Ala	Ala	Ile
915								920				925			
Gln	Gly	Ala	Leu	Asn	Gly	Pro	Asp	His	Pro	Pro	Lys	Asp	Leu	Arg	Ile
930						935						940			
Arg	Arg	Gln	His	Ser	Ser	Glu	Ser	Val	Ser	Ser	Ile	Asn	Ser	Ala	Thr
945						950				955				960	
Ser	His	Ser	Ser	Ile	Gly	Ser	Gly	Asn	Asp	Ala	Asp	Ser	Lys	Lys	
				965						970				975	

<210>	21
<211>	2583
<212>	DNA
<213>	mouse

<400> 21

gggatgaagg gaaaaagcca ccgtcaggca ttggaagatc gacagccagc agttctttttg 60  
gatacaagaa gccaaagtggg ttagggggctt ccactatgat taccagcagc ggtgccacca 120  
tcacaagcgg ttcagctaca ctgggggaaaa tccccaatc cgctgccatt ggtgggaagt 180

ccaatgcagg	aaggaaaacc	agcctggacg	gggtcccagaa	tcaagatgat	gttgtcctgc	240
acgtgagctc	gaagaccacc	ctccagtacc	gtagtttgcc	ccgcccttct	aagtccagca	300
ccagcggaat	ccctgggaga	ggtggccaca	ggtcgagcac	cagcagcatt	gattccaatg	360
tcagcagcaa	gtcagctggg	gccaccacct	ccaaactgag	agaaccgact	aagatcggct	420
cagggcgctc	gagtccagtc	actgtcaacc	aaacagacaa	agagaaggag	aaagtagcag	480
tgtcagattc	agagagcggt	tccttgtcag	gttcccccaa	atccagcccc	acctctgcca	540
gtgcctgtgg	gactcaaggg	ctcagacagc	cagggtccaa	atatccagat	attgcctcgc	600
ccacatttcg	aaggtaaggg	tatgtaaaga	gatgttggga	aaacataaaa	ggtagtatat	660
agcatgtatt	tattctgtac	gaaactattt	tcatgtattc	taaatattct	aagattctgt	720
atcttatact	tgtctaaaat	atagtgattt	tattttgctg	attgcacctg	ttgctagtgt	780
aaaagcattg	ctcattttaga	gagtggttag	cctttcagct	atacagccag	tgtgacacta	840
aaatacagat	accacttgta	gcgggcataa	aaccacatga	ctgactattc	atagaaataa	900
agtgatagct	tgtaaagata	tttagtgatt	tccacctctc	ctttccagaa	ttaaaaaaag	960
caaattgcat	agatctttat	aaacacattt	acttctagtg	tatgttatct	tgttgactct	1020
taatgaaatg	gcagttatga	atatagatga	tatattcttt	ctaacagttt	ataagagacc	1080
aatttataca	gtaccagatc	ttaacatagt	aacaataaca	gcaacaaaaa	caacccaaaa	1140
agctatcaaa	gtatggctcg	attgcagaat	ttgaaaacat	ttacatgttt	gacataggac	1200
aagaactcag	gagtgaggtg	actttttata	agtcttcatc	aatgtccttt	tacaggaacc	1260
aggaagcata	tctgatatat	gtgtcaggat	tatcacttta	ttaattatgt	gaaattctgt	1320
ttagaaatct	acctgatttt	aaatacttta	atatagtagg	ggtcaaaatt	agttaatgag	1380
ttaagacaag	ttgttaaata	atcctggctc	tgttttctca	tcttcaaaat	gatagagtat	1440
aatttatcac	ctcttgttaa	atatttcagg	tttgtgttta	ttctcttgat	aactttgac	1500
tcttagaaga	gtcttgaaga	atttacatta	agtaatctta	gaaacataac	tatttgagaa	1560
acagtagtca	aattttgtca	ttagaagtat	taactctgaa	gaatgatttg	aagtgacagt	1620
tcttagaaaag	aataaattat	agcttgtagc	aagagtaa	atcttcactg	cttgtgtgag	1680
agccaagagc	gccctcttgt	ggcccatcac	ctatgaaaca	atctctcata	ttcgccctag	1740
aaatcttcca	ctgcaggaaa	taatggattt	cattgcctct	gaattagtaa	ccattctgcc	1800
atttcttcac	accattttat	ttccatactt	gcataaattt	gattatgtca	tctgcttcac	1860
ttacaaaact	aaaatgtttt	ctgagctaaa	ctccagtagc	taacttagta	caaatgggtat	1920

tttttaaatca	ctgctataag	tatatatatatt	tgaatagctc	tggcaacgga	cggaaatccc	1980
tatgggtcttt	ccatgggaag	atacaaacca	atccataagt	tgtccagcaa	tatccaatat	2040
ttccagccca	gccagtcagg	cctcttaaac	attaccttac	atatttgaac	ctttccttaa	2100
atgtccccctt	tagacaatct	atttttttaa	aagatgaaaa	tccatttaag	catcatatat	2160
cgaatgcgta	gaagttgttt	cattataatg	gttctgcaga	taggtaatgc	caaaacggcc	2220
aaaatatattg	atcactagaa	gcgtaaaaagt	caagtacaat	catgttgact	ttttttccaa	2280
ggtgggttca	ctgctgcca	ccttggttcc	aggccagtgc	ttacttaaga	tatcgtaagt	2340
gattttttttt	taatttttaa	tttttttagta	gttggttaat	caaaagccag	tcatgtccacc	2400
ttcaggaaca	tagaggctgg	acgtgcttgg	cagctcacga	ctccaaagca	cacttggtctc	2460
tgtggactga	aaccctagga	aacgtggatg	tgagtctctt	ggaacaactc	aagttgttat	2520
ttgttttttct	tttaggttgt	tcggtgcca	ggcaggcggc	aaatctgcct	ccgcacctaa	2580
tac						2583

```
<210> 22
<211> 37
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> (2) .. (34)
```

<400> 22

c cgc ggg gct tcc atc ctt cct ttg act gat ttt taa 37  
Arg Gly Ala Ser Ile Leu Pro Leu Thr Asp Phe \*

```
<210> 23
<211> 6768
<212> DNA
<213> Homo sapiens
```

```
<220>
<221> CDS
<222> (1)..(5928)
```

<400> 23

att tta att tgt att ttc ccc gcc gcc ccg ccc ctt ttc ctc cga ccc 48



ctg tcc aag tcg gag cac tcg ctc ttc cag gcc aag ggc agc ccg gcg Leu Ser Lys Ser Glu His Ser Leu Phe Gln Ala Lys Gly Ser Pro Ala	768
ggc ggc gcc aag acc ccc ctg gct ccg ctc gcg ccc aac ctg gga aag Gly Gly Ala Lys Thr Pro Leu Ala Pro Leu Ala Pro Asn Leu Gly Lys	816
ccg agc cgg atc cct cga gga ccc tat gcg gag gtc aag ccg ctc agc Pro Ser Arg Ile Pro Arg Gly Pro Tyr Ala Glu Val Lys Pro Leu Ser	864
aag gcg cct gaa gcg gcc gtg agc gaa gat ggc aaa tcg gac gac gag Lys Ala Pro Glu Ala Ala Val Ser Glu Asp Gly Lys Ser Asp Asp Glu	912
ctg ctc tcc agc aag gcc aag gcg caa aag agc tct ggg cct gtc ccc Leu Leu Ser Ser Lys Ala Lys Ala Gln Lys Ser Ser Gly Pro Val Pro	960
tct gcc aag ggc cag gag gag cgc gcc ttc ctc aag gtg gac ccc gag Ser Ala Lys Gly Gln Glu Glu Arg Ala Phe Leu Lys Val Asp Pro Glu	1008
ctg gtg gtg acc gtg ctg gga gac ctg gag cag ctg ctc ttc agc cag Leu Val Val Thr Val Leu Gly Asp Leu Glu Gln Leu Leu Phe Ser Gln	1056
atg ctg gac cca gag tcc cag aga aag agg aca gtg cag aat gtc ctg Met Leu Asp Pro Glu Ser Gln Arg Lys Arg Thr Val Gln Asn Val Leu	1104
gat ctc cgg cag aac ctg gaa gag acc atg tcc agc ctg cga ggg tcc Asp Leu Arg Gln Asn Leu Glu Glu Thr Met Ser Ser Leu Arg Gly Ser	1152
cag gtg act cac agc tcc ctg gag atg acc tgc tac gac agc gat gat Gln Val Thr His Ser Ser Leu Glu Met Thr Cys Tyr Asp Ser Asp Asp	1200
gcc aac cca cgc agc gtg tcc agc ctc tcc aac cgc tcg tac cct ctg Ala Asn Pro Arg Ser Val Ser Ser Leu Ser Asn Arg Ser Tyr Pro Leu	1248
tca tgg cgc tat ggc cag tcc agt ccg cgg ctg cag gct ggt gac gcg Ser Trp Arg Tyr Gly Gln Ser Ser Pro Arg Leu Gln Ala Gly Asp Ala	1296
ccc tct gtg ggt ggg agc tgc cgc tcg gag ggg acg ccc gcc tgg tac Pro Ser Val Gly Gly Ser Cys Arg Ser Glu Gly Thr Pro Ala Trp Tyr	1344
atg cac ggc gaa cgg gcc cac tac tcc cac acc atg ccc atg cgc agc Met His Gly Glu Arg Ala His Tyr Ser His Thr Met Pro Met Arg Ser	1392

ccc	agc	aag	ctc	agc	cat	atc	tcc	cgc	ctg	gag	ctg	gtc	gaa	tcc	ctg	1440
Pro	Ser	Lys	Leu	Ser	His	Ile	Ser	Arg	Leu	Glu	Leu	Val	Glu	Ser	Leu	
gac	tcg	gat	gag	gtg	gac	ctc	aag	tcc	ggc	tac	atg	agc	gac	agt	gac	1488
Asp	Ser	Asp	Glu	Val	Asp	Leu	Lys	Ser	Gly	Tyr	Met	Ser	Asp	Ser	Asp	
ctc	atg	ggc	aag	acc	atg	acg	gag	gat	gat	gac	atc	act	acc	ggc	tgg	1536
Leu	Met	Gly	Lys	Thr	Met	Thr	Glu	Asp	Asp	Asp	Ile	Thr	Thr	Gly	Trp	
gat	gaa	agc	agc	tcc	atc	agt	agt	gga	ctc	agc	gat	gcc	tca	gac	aat	1584
Asp	Glu	Ser	Ser	Ser	Ile	Ser	Ser	Gly	Leu	Ser	Asp	Ala	Ser	Asp	Asn	
ctc	agt	tca	gaa	gaa	ttc	aat	gcc	agc	tcc	tca	ctc	aac	tcc	ctc	cca	1632
Leu	Ser	Ser	Glu	Glu	Phe	Asn	Ala	Ser	Ser	Ser	Leu	Asn	Ser	Leu	Pro	
agt	act	ccc	act	gct	tct	cgc	agg	aac	tca	aca	ata	gtg	cta	cgc	aca	1680
Ser	Thr	Pro	Thr	Ala	Ser	Arg	Arg	Asn	Ser	Thr	Ile	Val	Leu	Arg	Thr	
gac	tca	gag	aag	cgc	tca	ctg	gca	gaa	agt	ggg	ctg	agc	tgg	ttt	agt	1728
Asp	Ser	Glu	Lys	Arg	Ser	Leu	Ala	Glu	Ser	Gly	Leu	Ser	Trp	Phe	Ser	
gaa	tca	gag	gag	aaa	gcc	cct	aaa	aaa	ctg	gag	tac	gac	agt	ggg	agc	1776
Glu	Ser	Glu	Glu	Lys	Ala	Pro	Lys	Lys	Leu	Glu	Tyr	Asp	Ser	Gly	Ser	
ctg	aag	atg	gaa	cct	ggg	act	tct	aag	tgg	cgg	agg	gag	cgg	cct	gag	1824
Leu	Lys	Met	Glu	Pro	Gly	Thr	Ser	Lys	Trp	Arg	Arg	Glu	Arg	Pro	Glu	
agc	tgt	gat	gat	tca	tcc	aag	ggg	gga	gaa	ctg	aaa	aag	ccc	atc	agc	1872
Ser	Cys	Asp	Asp	Ser	Ser	Lys	Gly	Gly	Glu	Leu	Lys	Lys	Pro	Ile	Ser	
ctg	ggc	cac	cct	ggg	tcc	ctg	aag	aag	ggc	aag	acc	cca	cct	gtg	gct	1920
Leu	Gly	His	Pro	Gly	Ser	Leu	Lys	Lys	Gly	Lys	Thr	Pro	Pro	Val	Ala	
gta	act	tcc	ccc	atc	act	cac	aca	gcc	cag	agt	gcc	ctc	aaa	gtc	gca	1968
Val	Thr	Ser	Pro	Ile	Thr	His	Thr	Ala	Gln	Ser	Ala	Leu	Lys	Val	Ala	
ggc	aaa	cct	gag	ggc	aaa	gct	aca	gac	aag	ggg	aag	ctt	gca	gtg	aag	2016
Gly	Lys	Pro	Glu	Gly	Lys	Ala	Thr	Asp	Lys	Gly	Lys	Leu	Ala	Val	Lys	
aat	act	ggg	ctc	caa	cgc	tcc	tcc	tct	gat	gct	ggg	cgg	gac	cgc	ctg	2064
Asn	Thr	Gly	Leu	Gln	Arg	Ser	Ser	Ser	Asp	Ala	Gly	Arg	Asp	Arg	Leu	



09034549 060502

cat gct aca agc tca gca tct ggg ggc cct ctc cct tcc tgc ttc acc	2832
His Ala Thr Ser Ser Ala Ser Gly Gly Pro Leu Pro Ser Cys Phe Thr	
ccc agt ccg gca ccc atc ctc aat att aac tca gcc agc ttc tcc cag	2880
Pro Ser Pro Ala Pro Ile Leu Asn Ile Asn Ser Ala Ser Phe Ser Gln	
ggc ctg gag cta atg agt ggt ttc agt gtg cca aaa gag acc cgc atg	2928
Gly Leu Glu Leu Met Ser Gly Phe Ser Val Pro Lys Glu Thr Arg Met	
tac ccc aaa ctc tca ggc ctg cac agg agc atg gag tcc ctc cag atg	2976
Tyr Pro Lys Leu Ser Gly Leu His Arg Ser Met Glu Ser Leu Gln Met	
cca atg agc ctc ccc agt gcc ttc ccc agc agt act ccc gtc ccc acc	3024
Pro Met Ser Leu Pro Ser Ala Phe Pro Ser Ser Thr Pro Val Pro Thr	
cca cct gct ccc cct gct gct ccc aca gaa gaa gag acg gaa gag ctg	3072
Pro Pro Ala Pro Pro Ala Ala Pro Thr Glu Glu Glu Thr Glu Glu Leu	
act tgg agt gga agc ccc aga gct ggg caa ctg gac agt aat cag cgg	3120
Thr Trp Ser Gly Ser Pro Arg Ala Gly Gln Leu Asp Ser Asn Gln Arg	
gat cgg aac act ctt ccc aag aaa ggg ctc agg tac cag ctt cag tcc	3168
Asp Arg Asn Thr Leu Pro Lys Lys Gly Leu Arg Tyr Gln Leu Gln Ser	
cag gag gag acc aag gag agg cga cat tcc cat acc att ggt ggg ctg	3216
Gln Glu Glu Thr Lys Glu Arg Arg His Ser His Thr Ile Gly Gly Leu	
cct gaa tcc gat gac cag tca gag ctg cct tct ccc cct gca ctt ccc	3264
Pro Glu Ser Asp Asp Gln Ser Glu Leu Pro Ser Pro Pro Ala Leu Pro	
atg tct ctg agt gca aag ggc caa ctt acc aac ata gtg agt ccc act	3312
Met Ser Leu Ser Ala Lys Gly Gln Leu Thr Asn Ile Val Ser Pro Thr	
gcg gcc acc acg cca aga atc acc cgc tcc aac agc atc ccc acc cac	3360
Ala Ala Thr Thr Pro Arg Ile Thr Arg Ser Asn Ser Ile Pro Thr His	
gag gcg gcc ttc gag ctg tac agc ggc tcc caa atg ggg agc acc ctg	3408
Glu Ala Ala Phe Glu Leu Tyr Ser Gly Ser Gln Met Gly Ser Thr Leu	
tcc ctg gcc gag aga ccc aag gga atg att cgg tca gga tcc ttc cga	3456
Ser Leu Ala Glu Arg Pro Lys Gly Met Ile Arg Ser Gly Ser Phe Arg	
gac ccc acg gac gat gtt cac ggc tca gtg ctg tcc ctg gcc tcc agt	3504
Asp Pro Thr Asp Asp Val His Gly Ser Val Leu Ser Leu Ala Ser Ser	
gcc tcc tcc acc tac tcc tca gct gag gag agg atg caa tct gag caa	3552
Ala Ser Ser Thr Tyr Ser Ser Ala Glu Glu Arg Met Gln Ser Glu Gln	
atc cgg aag ctt cgt agg gaa ctg gaa tca tcc cag gaa aaa gtg gcc	3600
Ile Arg Lys Leu Arg Arg Glu Leu Glu Ser Ser Gln Glu Lys Val Ala	
acc ttg acg tct cag ctt tct gcc aat gct aat ctg gtg gct gct ttt	3648
Thr Leu Thr Ser Gln Leu Ser Ala Asn Ala Asn Leu Val Ala Ala Phe	
gag cag agc ctg gtg aat atg aca tcc cgc ctg cga cac ctg gca gag	3696
Glu Gln Ser Leu Val Asn Met Thr Ser Arg Leu Arg His Leu Ala Glu	

acg gcc gag gag aag gac act gag ctg ctg gat ttg cga gaa acc ata Thr Ala Glu Glu Lys Asp Thr Glu Leu Leu Asp Leu Arg Glu Thr Ile	3744
gac ttt ctg aag aaa aag aac tct gag gcc cag gca gtc att cag gga Asp Phe Leu Lys Lys Lys Asn Ser Glu Ala Gln Ala Val Ile Gln Gly	3792
gcc ctt aat gcc tca gaa acc aca ccc aaa gaa ctt cgg atc aag aga Ala Leu Asn Ala Ser Glu Thr Thr Pro Lys Glu Leu Arg Ile Lys Arg	3840
caa aac tcc tca gat agc atc tca agc ctc aac agc atc act agc cat Gln Asn Ser Ser Asp Ser Ile Ser Ser Leu Asn Ser Ile Thr Ser His	3888
tcc agc atc ggc agc agc aag gat gct gat gcg aaa aag aag aaa aaa Ser Ser Ile Gly Ser Ser Lys Asp Ala Asp Ala Lys Lys Lys Lys Lys	3936
aag agt tgg ctt cga agt tcc ttc aac aaa gcg ttc agt ata aaa aag Lys Ser Trp Leu Arg Ser Ser Phe Asn Lys Ala Phe Ser Ile Lys Lys	3984
ggg ccc aag tca gct tcc tca tac tcg gat ata gag gag att gct aca Gly Pro Lys Ser Ala Ser Ser Tyr Ser Asp Ile Glu Glu Ile Ala Thr	4032
ccc gac tct tca gcc ccc tca tcc ccc aaa cta cag cat ggt tct aca Pro Asp Ser Ser Ala Pro Ser Ser Pro Lys Leu Gln His Gly Ser Thr	4080
gag act gct tca ccc tcc atc aag tcc tcc acc tcg tcc tcc gtg ggc Glu Thr Ala Ser Pro Ser Ile Lys Ser Ser Thr Ser Ser Ser Val Gly	4128
act gat gtc acc gag ggc cct gct cac cca gcc ccc cac act agg ctg Thr Asp Val Thr Glu Gly Pro Ala His Pro Ala Pro His Thr Arg Leu	4176
ttc cat gca aat gag gag gag gag cca gag aag aag gag gta tcg gag Phe His Ala Asn Glu Glu Glu Glu Pro Glu Lys Lys Glu Val Ser Glu	4224
ctg cgc tct gag cta tgg gag aag gaa atg aag ctt aca gac atc cgc Leu Arg Ser Glu Leu Trp Glu Lys Glu Met Lys Leu Thr Asp Ile Arg	4272
ttg gag gcc ctc aac tct gcc cac caa ctg gat cag ctt cgg gag acc Leu Glu Ala Leu Asn Ser Ala His Gln Leu Asp Gln Leu Arg Glu Thr	4320
atg cac aac atg cag ttg gag gtg gac ctg ctg gaa gca gag aat gac Met His Asn Met Gln Leu Glu Val Asp Leu Leu Glu Ala Glu Asn Asp	4368
cga ctg aag gta gcc cca ggc ccc tca tca ggc tcc act cca ggg cag Arg Leu Lys Val Ala Pro Gly Pro Ser Ser Gly Ser Thr Pro Gly Gln	4416
gtc cct gga tca tct gca tta tct tcc cca cgc cgc tcc cta ggc ctg Val Pro Gly Ser Ser Ala Leu Ser Ser Pro Arg Arg Ser Leu Gly Leu	4464
gca ctc acc cat tcc ttc ggc ccc agt ctt gca gac aca gac ctg tca Ala Leu Thr His Ser Phe Gly Pro Ser Leu Ala Asp Thr Asp Leu Ser	4512
ccc atg gat ggc atc agt act tgt ggt cca aag gag gaa gtg acc ctc Pro Met Asp Gly Ile Ser Thr Cys Gly Pro Lys Glu Glu Val Thr Leu	4560
cgg gtg gtg gtg agg atg ccc ccg cag cac atc atc aaa ggg gac ttg Arg Val Val Val Arg Met Pro Pro Gln His Ile Ile Lys Gly Asp Leu	4608

aag cag cag gaa ttc ttc ctg ggc tgt agc aag gtc agt gga aaa gtt Lys Gln Gln Glu Phe Phe Leu Gly Cys Ser Lys Val Ser Gly Lys Val	4656
gac tgg aag atg ctg gat gaa gct gtt ttc caa gtg ttc aag gac tat Asp Trp Lys Met Leu Asp Glu Ala Val Phe Gln Val Phe Lys Asp Tyr	4704
att tct aaa atg gac cca gcc tct acc ctg gga cta agc act gag tcc Ile Ser Lys Met Asp Pro Ala Ser Thr Leu Gly Leu Ser Thr Glu Ser	4752
atc cat ggc tac agc atc agc cac gtg aaa cga gtg ttg gat gca gag Ile His Gly Tyr Ser Ile Ser His Val Lys Arg Val Leu Asp Ala Glu	4800
ccc ccc gag atg cct cct tgc cgt cga ggt gtc aat aac ata tca gtc Pro Pro Glu Met Pro Pro Cys Arg Arg Gly Val Asn Asn Ile Ser Val	4848
tcc ctc aaa ggt ctg aag gag aaa tgc gtc gac agc ctg gtg ttc gag Ser Leu Lys Gly Leu Lys Glu Lys Cys Val Asp Ser Leu Val Phe Glu	4896
acg ctg atc ccc aag ccg atg atg cag cac tac ata agc ctc ctg ctg Thr Leu Ile Pro Lys Pro Met Met Gln His Tyr Ile Ser Leu Leu Leu	4944
aag cac cgg cgc ctc gtc ctc tcg ggc ccc agc ggc acg ggc aag acc Lys His Arg Arg Leu Val Leu Ser Gly Pro Ser Gly Thr Gly Lys Thr	4992
tac ctg acc aat cgc ttg gcc gag tac ctg gtg gag cgc tct ggc cgt Tyr Leu Thr Asn Arg Leu Ala Glu Tyr Leu Val Glu Arg Ser Gly Arg	5040
gag gtc aca gag ggc atc gtc agc acc ttc aac atg cac cag cag tct Glu Val Thr Glu Gly Ile Val Ser Thr Phe Asn Met His Gln Gln Ser	5088
tgc aag gat ctg caa ctg tat ctt tcc aac cta gcc aac cag ata gac Cys Lys Asp Leu Gln Leu Tyr Leu Ser Asn Leu Ala Asn Gln Ile Asp	5136
cgg gaa aca gga att ggg gat gtg ccc ctg gtg att cta ttg gat gac Arg Glu Thr Gly Ile Gly Asp Val Pro Leu Val Ile Leu Leu Asp Asp	5184
ctg agt gaa gca ggc tcc atc agt gag ttg gtc aat ggg gcc ctc acc Leu Ser Glu Ala Gly Ser Ile Ser Glu Leu Val Asn Gly Ala Leu Thr	5232
tgc aag tat cat aaa tgt ccc tat att ata ggt acc acc aat cag cct Cys Lys Tyr His Lys Cys Pro Tyr Ile Ile Gly Thr Thr Asn Gln Pro	5280
gta aaa atg aca ccc aac cat ggc ttt cac ttg agc ttc agg atg ttg Val Lys Met Thr Pro Asn His Gly Phe His Leu Ser Phe Arg Met Leu	5328
acc ttc tcc aac aac gtg gag cca gcc aat ggc ttc ctg gtt cgt tac Thr Phe Ser Asn Asn Val Glu Pro Ala Asn Gly Phe Leu Val Arg Tyr	5376
ctg agg agg aag ctg gta gag tca gac agc gac atc aat gcc aac aag Leu Arg Arg Lys Leu Val Glu Ser Asp Ser Asp Ile Asn Ala Asn Lys	5424
gaa gag ctg ctt cgg gtg ctc gac tgg gta ccc aag ctg tgg tat cat Glu Glu Leu Leu Arg Val Leu Asp Trp Val Pro Lys Leu Trp Tyr His	5472
ctc cac acc ttc ctt gag aag cac agc acc tca gac ttc ctc atc ggc	5520

Leu His Thr Phe Leu Glu Lys His Ser Thr Ser Asp Phe Leu Ile Gly  
 cct tgc ttc ttt ctg tgc tgt ccc att ggc att gag gac ttc cgg acc 5568  
 Pro Cys Phe Phe Leu Ser Cys Pro Ile Gly Ile Glu Asp Phe Arg Thr  
 tgg ttc att gac ctg tgg aac aac tct atc att ccc tat cta cag gaa 5616  
 Trp Phe Ile Asp Leu Trp Asn Asn Ser Ile Ile Pro Tyr Leu Gln Glu  
 gga gcc aag gat ggg ata aag gtc cat gga cag aaa gct gct tgg gag 5664  
 Gly Ala Lys Asp Gly Ile Lys Val His Gly Gln Lys Ala Ala Trp Glu  
 gac cca gtg gaa tgg gtc cgg gac aca ctt ccc tgg cca tca gcc caa 5712  
 Asp Pro Val Glu Trp Val Arg Asp Thr Leu Pro Trp Pro Ser Ala Gln  
 caa gac caa tca aag ctg tac cac ctg ccc cca ccc acc gtg ggc cct 5760  
 Gln Asp Gln Ser Lys Leu Tyr His Leu Pro Pro Pro Thr Val Gly Pro  
 cac agc att gcc tca cct ccc gag gat agg aca gtc aaa gac agc acc 5808  
 His Ser Ile Ala Ser Pro Pro Glu Asp Arg Thr Val Lys Asp Ser Thr  
 cca agt tct ctg gac tca gat cct ctg atg gcc atg ctg ctg aaa ctt 5856  
 Pro Ser Ser Leu Asp Ser Asp Pro Leu Met Ala Met Leu Leu Lys Leu  
 caa gaa gct gcc aac tac att gag tct cca gat cga gaa acc atc ctg 5904  
 Gln Glu Ala Ala Asn Tyr Ile Glu Ser Pro Asp Arg Glu Thr Ile Leu  
 gac ccc aac ctt cag gca aca ctt taagggttcg gcaatcactg 5948  
 Asp Pro Asn Leu Gln Ala Thr Leu  
 tcacccccgg acagcagaac gctggcatca gctatcttag ctctctctct cccctctct 6008  
 ctttcagagc actgggtctc cagccccagg aggagaacag gagggaggag gagatgaaag 6068  
 aggagggaca gggtcttggg gctgtacctt tgagaacttc ctaggaagga atggtggggg 6128  
 ggcgtttggg aacttgtgcc ccctaaacac atttactggc ctctctaat gactttgggg 6188  
 aaaagatgat tctgggtctt tcccttgact tcttggttca attacaaact cctgggcttt 6248  
 ctggggaggg gttcagaaaa catcaaaaca ctgcagcagt tctaaatga ttctcacaag 6308  
 caaccctgag agagacagtc ttgtgaggga gatctggggg aggcaggaag ctctcagat 6368  
 tttctcacag acccttccca attccatcac cactgccaac aactcctccc ccagagatct 6428  
 ggctggagcc cagaaaaaga agcatgtggg ttaaaaaatg tttaaataaa tctgtaaaag 6488  
 gtaaaaaatga aaaacaaaaa caagcaaaaca aacaaaaaac aatggaaaaag atgaagctgg 6548  
 agagagagga accagttgcc aaggtagaga gctgcccgt cctgccctct ggatgacata 6608  
 ggggacatca acaagacggc tgccaacctg agaagtcacc aaaccacaaa aataacctta 6668  
 cagccttcag ggaaagacta ccagctctgt ctttctaccc tctaatttaa caatgcataa 6728  
 gagtcaataa accctacttt tttaaaaaaa aaaaaaaaag 6768



165										170					175				
Arg	Val	Pro	Gly 180	Gly	Pro	Pro	Ala	Ser 185	Asn	Leu	Arg	Lys	Gln 190	Lys	Ser				
Leu	Thr	Asn 195	Leu	Ser	Phe	Leu	Thr 200	Asp	Ser	Glu	Lys	Lys 205	Leu	Gln	Leu				
Tyr	Glu 210	Pro	Glu	Trp	Ser	Asp 215	Asp	Met	Ala	Lys	Ala 220	Pro	Lys	Gly	Leu				
Gly 225	Lys	Val	Gly	Ser	Lys 230	Gly	Arg	Glu	Ala	Pro 235	Leu	Met	Ser	Lys	Thr 240				
Leu	Ser	Lys	Ser	Glu 245	His	Ser	Leu	Phe	Gln 250	Ala	Lys	Gly	Ser	Pro 255	Ala				
Gly	Gly	Ala	Lys 260	Thr	Pro	Leu	Ala	Pro 265	Leu	Ala	Pro	Asn 270	Leu	Gly	Lys				
Pro	Ser	Arg 275	Ile	Pro	Arg	Gly	Pro 280	Tyr	Ala	Glu	Val	Lys 285	Pro	Leu	Ser				
Lys	Ala 290	Pro	Glu	Ala	Ala	Val 295	Ser	Glu	Asp	Gly	Lys 300	Ser	Asp	Asp	Glu				
Leu 305	Leu	Ser	Ser	Lys 310	Ala	Lys	Ala	Gln	Lys 315	Ser	Ser	Gly	Pro	Val 320	Pro				
Ser	Ala	Lys	Gly	Gln 325	Glu	Glu	Arg	Ala	Phe 330	Leu	Lys	Val	Asp	Pro 335	Glu				
Leu	Val	Val	Thr 340	Val	Leu	Gly	Asp 345	Leu	Glu	Gln	Leu	Leu	Phe 350	Ser	Gln				
Met	Leu	Asp 355	Pro	Glu	Ser	Gln	Arg 360	Lys	Arg	Thr	Val	Gln 365	Asn	Val	Leu				
Asp	Leu 370	Arg	Gln	Asn	Leu	Glu 375	Glu	Thr	Met	Ser	Ser 380	Leu	Arg	Gly	Ser				
Gln 385	Val	Thr	His	Ser 390	Ser	Leu	Glu	Met	Thr	Cys 395	Tyr	Asp	Ser	Asp	Asp 400				
Ala	Asn	Pro	Arg	Ser 405	Val	Ser	Ser	Leu	Ser 410	Asn	Arg	Ser	Tyr	Pro 415	Leu				
Ser	Trp	Arg	Tyr 420	Gly	Gln	Ser	Ser	Pro	Arg 425	Leu	Gln	Ala	Gly 430	Asp	Ala				
Pro	Ser	Val	Gly 435	Gly	Ser	Cys	Arg 440	Ser	Glu	Gly	Thr	Pro 445	Ala	Trp	Tyr				
Met	His 450	Gly	Glu	Arg	Ala	His 455	Tyr	Ser	His	Thr	Met 460	Pro	Met	Arg	Ser				

Pro Ser Lys Leu Ser His Ile Ser Arg Leu Glu Leu Val Glu Ser Leu  
 465 470 475 480  
 Asp Ser Asp Glu Val Asp Leu Lys Ser Gly Tyr Met Ser Asp Ser Asp  
 485 490 495  
 Leu Met Gly Lys Thr Met Thr Glu Asp Asp Ile Thr Thr Gly Trp  
 500 505 510  
 Asp Glu Ser Ser Ser Ile Ser Ser Gly Leu Ser Asp Ala Ser Asp Asn  
 515 520 525  
 Leu Ser Ser Glu Glu Phe Asn Ala Ser Ser Ser Leu Asn Ser Leu Pro  
 530 535 540  
 Ser Thr Pro Thr Ala Ser Arg Arg Asn Ser Thr Ile Val Leu Arg Thr  
 545 550 555 560  
 Asp Ser Glu Lys Arg Ser Leu Ala Glu Ser Gly Leu Ser Trp Phe Ser  
 565 570 575  
 Glu Ser Glu Glu Lys Ala Pro Lys Lys Leu Glu Tyr Asp Ser Gly Ser  
 580 585 590  
 Leu Lys Met Glu Pro Gly Thr Ser Lys Trp Arg Arg Glu Arg Pro Glu  
 595 600 605  
 Ser Cys Asp Asp Ser Ser Lys Gly Gly Glu Leu Lys Lys Pro Ile Ser  
 610 615 620  
 Leu Gly His Pro Gly Ser Leu Lys Lys Gly Lys Thr Pro Pro Val Ala  
 625 630 635 640  
 Val Thr Ser Pro Ile Thr His Thr Ala Gln Ser Ala Leu Lys Val Ala  
 645 650 655  
 Gly Lys Pro Glu Gly Lys Ala Thr Asp Lys Gly Lys Leu Ala Val Lys  
 660 665 670  
 Asn Thr Gly Leu Gln Arg Ser Ser Ser Asp Ala Gly Arg Asp Arg Leu  
 675 680 685  
 Ser Asp Ala Lys Lys Pro Pro Ser Gly Ile Ala Arg Pro Ser Thr Ser  
 690 695 700  
 Gly Ser Phe Gly Tyr Lys Lys Pro Pro Pro Ala Thr Gly Thr Ala Thr  
 705 710 715 720  
 Val Met Gln Thr Gly Gly Ser Ala Thr Leu Ser Lys Ile Gln Lys Ser  
 725 730 735  
 Ser Gly Ile Pro Val Lys Pro Val Asn Gly Arg Lys Thr Ser Leu Asp  
 740 745 750  
 Val Ser Asn Ser Ala Glu Pro Gly Phe Leu Ala Pro Gly Ala Arg Ser  
 755 760 765





Pro Glu Ser Asp Asp Gln Ser Glu Leu Pro Ser Pro Pro Ala Leu Pro  
 1075 1080 1085  
 Met Ser Leu Ser Ala Lys Gly Gln Leu Thr Asn Ile Val Ser Pro Thr  
 1090 1095 1100  
 Ala Ala Thr Thr Pro Arg Ile Thr Arg Ser Asn Ser Ile Pro Thr His  
 1105 1110 1115 1120  
 Glu Ala Ala Phe Glu Leu Tyr Ser Gly Ser Gln Met Gly Ser Thr Leu  
 1125 1130 1135  
 Ser Leu Ala Glu Arg Pro Lys Gly Met Ile Arg Ser Gly Ser Phe Arg  
 1140 1145 1150  
 Asp Pro Thr Asp Asp Val His Gly Ser Val Leu Ser Leu Ala Ser Ser  
 1155 1160 1165  
 Ala Ser Ser Thr Tyr Ser Ser Ala Glu Glu Arg Met Gln Ser Glu Gln  
 1170 1175 1180  
 Ile Arg Lys Leu Arg Arg Glu Leu Glu Ser Ser Gln Glu Lys Val Ala  
 1185 1190 1195 1200  
 Thr Leu Thr Ser Gln Leu Ser Ala Asn Ala Asn Leu Val Ala Ala Phe  
 1205 1210 1215  
 Glu Gln Ser Leu Val Asn Met Thr Ser Arg Leu Arg His Leu Ala Glu  
 1220 1225 1230  
 Thr Ala Glu Glu Lys Asp Thr Glu Leu Leu Asp Leu Arg Glu Thr Ile  
 1235 1240 1245  
 Asp Phe Leu Lys Lys Lys Asn Ser Glu Ala Gln Ala Val Ile Gln Gly  
 1250 1255 1260  
 Ala Leu Asn Ala Ser Glu Thr Thr Pro Lys Glu Leu Arg Ile Lys Arg  
 1265 1270 1275 1280  
 Gln Asn Ser Ser Asp Ser Ile Ser Ser Leu Asn Ser Ile Thr Ser His  
 1285 1290 1295  
 Ser Ser Ile Gly Ser Ser Lys Asp Ala Asp Ala Lys Lys Lys Lys Lys  
 1300 1305 1310  
 Lys Ser Trp Leu Arg Ser Ser Phe Asn Lys Ala Phe Ser Ile Lys Lys  
 1315 1320 1325  
 Gly Pro Lys Ser Ala Ser Ser Tyr Ser Asp Ile Glu Glu Ile Ala Thr  
 1330 1335 1340  
 Pro Asp Ser Ser Ala Pro Ser Ser Pro Lys Leu Gln His Gly Ser Thr  
 1345 1350 1355 1360  
 Glu Thr Ala Ser Pro Ser Ile Lys Ser Ser Thr Ser Ser Ser Val Gly







cag gtg gcc ggg gcc ccc tcc cag tgc cag gct ggc acc cct cag cag Gln Val Ala Gly Ala Pro Ser Gln Cys Gln Ala Gly Thr Pro Gln Gln	526
cag gtg cca gtc act ccc caa gcc ccg tgc cag cct cac cag cca gcg Gln Val Pro Val Thr Pro Gln Ala Pro Cys Gln Pro His Gln Pro Ala	574
cca cat cag cag tca aaa gca caa gct gaa atg cag tcc aga ctt cca Pro His Gln Gln Ser Lys Ala Gln Ala Glu Met Gln Ser Arg Leu Pro	622
ggt cct acc gcg agg gta tcc gct gca ggc agc gag gcc aaa aca cgc Gly Pro Thr Ala Arg Val Ser Ala Ala Gly Ser Glu Ala Lys Thr Arg	670
gga ggg tca act act gct aac aac cga cgc agc cag agc ttt aac aac Gly Gly Ser Thr Thr Ala Asn Asn Arg Arg Ser Gln Ser Phe Asn Asn	718
tat gat aaa tcc aaa cca gtc acc tcc cca ccc cca ccg cca agc agc Tyr Asp Lys Ser Lys Pro Val Thr Ser Pro Pro Pro Pro Pro Ser Ser	766
cac gag aaa gag cct ttg gca agt tca gcc tcc tcc cac ccc gga atg His Glu Lys Glu Pro Leu Ala Ser Ser Ala Ser Ser His Pro Gly Met	814
agt gac aat gca cct gct tcc ttg gag agc ggc agc agc tcc acc cct Ser Asp Asn Ala Pro Ala Ser Leu Glu Ser Gly Ser Ser Ser Thr Pro	862
act aat tgc agt acc tcc tcg gcc atc ccg cag ccc ggt gca gcc acc Thr Asn Cys Ser Thr Ser Ser Ala Ile Pro Gln Pro Gly Ala Ala Thr	910
aag cct tgg cgc agc aaa tcc ctc agc gtg aag cac agt gcc acg gta Lys Pro Trp Arg Ser Lys Ser Leu Ser Val Lys His Ser Ala Thr Val	958
tcc atg ctc tcg gtc aag cct cct ggg cct gag gcc ccc agg ccc aca Ser Met Leu Ser Val Lys Pro Pro Gly Pro Glu Ala Pro Arg Pro Thr	1006
cct gaa gcc atg aag ccg gcc ccc aac aat cag aag tcc atg ctg gaa Pro Glu Ala Met Lys Pro Ala Pro Asn Asn Gln Lys Ser Met Leu Glu	1054
aag ctg aaa ctt ttc aac agt aaa ggg ggc tca aag gca ggt gag ggg Lys Leu Lys Leu Phe Asn Ser Lys Gly Gly Ser Lys Ala Gly Glu Gly	1102
ccg ggg tcc cgg gac aca agc tgt gag cgg ctg gag act ctg ccc agc Pro Gly Ser Arg Asp Thr Ser Cys Glu Arg Leu Glu Thr Leu Pro Ser	1150









Ser Ser Ser Arg Thr Pro Thr Ala Asn Ala Asn Ser Phe Gly Phe Lys  
 aag cag agt ggt tcc gcc acc ggc ctg gcc atg atc aca gcc agc ggg 3262  
 Lys Gln Ser Gly Ser Ala Thr Gly Leu Ala Met Ile Thr Ala Ser Gly  
 gtg act gtc acc agc agg tca gcc aca ctg ggc aaa atc cca aag tca 3310  
 Val Thr Val Thr Ser Arg Ser Ala Thr Leu Gly Lys Ile Pro Lys Ser  
 tct gca ctc gtc agt cgg tct gct ggt cgg aag tca agt atg gat ggg 3358  
 Ser Ala Leu Val Ser Arg Ser Ala Gly Arg Lys Ser Ser Met Asp Gly  
 gct cag aat cag gat gac ggg tat cta gcc cta agc tcc cgg aca aac 3406  
 Ala Gln Asn Gln Asp Asp Gly Tyr Leu Ala Leu Ser Ser Arg Thr Asn  
 ctt cag tac cgg agt ttg ccg agg ccc agt aag tcc aac agc cgg aac 3454  
 Leu Gln Tyr Arg Ser Leu Pro Arg Pro Ser Lys Ser Asn Ser Arg Asn  
 ggg gct ggg aac agg tct agc acc agc agc ata gat tcc aac att agc 3502  
 Gly Ala Gly Asn Arg Ser Ser Thr Ser Ser Ile Asp Ser Asn Ile Ser  
 agc aag tcc gca ggc ctg cca gtg ccc aaa ctg agg gag cct tcc aaa 3550  
 Ser Lys Ser Ala Gly Leu Pro Val Pro Lys Leu Arg Glu Pro Ser Lys  
 aca gcc cta ggc agc tct cta cca ggt ctg gtc aac caa aca gac aag 3598  
 Thr Ala Leu Gly Ser Ser Leu Pro Gly Leu Val Asn Gln Thr Asp Lys  
 gag aaa ggc atc tca tca gac aac gag agt gtg gct tcc tgt aac tcg 3646  
 Glu Lys Gly Ile Ser Ser Asp Asn Glu Ser Val Ala Ser Cys Asn Ser  
 gtg aaa gtg aat ccg gca gcc cag cct gtg tcc agt ccg gct cag acc 3694  
 Val Lys Val Asn Pro Ala Ala Gln Pro Val Ser Ser Pro Ala Gln Thr  
 agt ctc cag cct gga gcc aag tac cca gat gtg gcc tct ccc aca ctc 3742  
 Ser Leu Gln Pro Gly Ala Lys Tyr Pro Asp Val Ala Ser Pro Thr Leu  
 cgc aga ctc ttt ggt ggg aag cct acc aag caa gtg ccc atc gcc aca 3790  
 Arg Arg Leu Phe Gly Gly Lys Pro Thr Lys Gln Val Pro Ile Ala Thr  
 gct gaa aac atg aaa aat tcg gtg gtc atc tcc aat cct cat gcc acc 3838  
 Ala Glu Asn Met Lys Asn Ser Val Val Ile Ser Asn Pro His Ala Thr  
 atg act cag caa ggt aac cta gac tcc ccg tca ggc agt ggc gtc ctg 3886  
 Met Thr Gln Gln Gly Asn Leu Asp Ser Pro Ser Gly Ser Gly Val Leu

agc agt ggg agc agc agt cct ctc tac agc aag aat gtg gac ctc aac Ser Ser Gly Ser Ser Ser Pro Leu Tyr Ser Lys Asn Val Asp Leu Asn	3934
cag tct ccg cta gcc tcc agc ccc agc tca gcc cac tcg gcc cct tcc Gln Ser Pro Leu Ala Ser Ser Pro Ser Ser Ala His Ser Ala Pro Ser	3982
aac agc ctc acc tgg ggc acc aac gcc agc agc tcc tcc gca gtt agc Asn Ser Leu Thr Trp Gly Thr Asn Ala Ser Ser Ser Ser Ala Val Ser	4030
aag gat ggc ctg ggc ttt cag tct gtc agc agc ctc cac acc agc tgt Lys Asp Gly Leu Gly Phe Gln Ser Val Ser Ser Leu His Thr Ser Cys	4078
gag tcc atc gac atc tcc ctc agc agt gga ggg gtc ccc agc cac aat Glu Ser Ile Asp Ile Ser Leu Ser Ser Gly Gly Val Pro Ser His Asn	4126
tct tcc act ggc ctc atc gcc tcc tcc aag gac gac tcc ttg act ccc Ser Ser Thr Gly Leu Ile Ala Ser Ser Lys Asp Asp Ser Leu Thr Pro	4174
ttt gtc aga act aac agt gtg aag acc aca ctg tca gaa agc cct ctc Phe Val Arg Thr Asn Ser Val Lys Thr Thr Leu Ser Glu Ser Pro Leu	4222
tct tcc cct gct gct agc cct aag ttc tgc aga agt act ctg ccc agg Ser Ser Pro Ala Ala Ser Pro Lys Phe Cys Arg Ser Thr Leu Pro Arg	4270
aaa cag gac agt gac ccg cac ctt gat agg aac act ttg cct aag aaa Lys Gln Asp Ser Asp Pro His Leu Asp Arg Asn Thr Leu Pro Lys Lys	4318
gga ctc agg tat act ccc acc tcc cag ctt cgc acg caa gaa gat gca Gly Leu Arg Tyr Thr Pro Thr Ser Gln Leu Arg Thr Gln Glu Asp Ala	4366
aaa gaa tgg tta cgg tcc cat tct gca gga ggc ctt cag gac acc gct Lys Glu Trp Leu Arg Ser His Ser Ala Gly Gly Leu Gln Asp Thr Ala	4414
gcc aat tcc ccc ttt tcc tct ggc tcc agc gtg act tct ccc tcc gga Ala Asn Ser Pro Phe Ser Ser Gly Ser Ser Val Thr Ser Pro Ser Gly	4462
aca aga ttc aac ttt tcc cag ctt gcg agt ccc acc act gtc acc cag Thr Arg Phe Asn Phe Ser Gln Leu Ala Ser Pro Thr Thr Val Thr Gln	4510
atg agc ttg tcc aac ccg acc atg ctg agg act cac agc ctc tcc aat Met Ser Leu Ser Asn Pro Thr Met Leu Arg Thr His Ser Leu Ser Asn	4558

gct	gat	ggg	cag	tat	gat	cca	tac	act	gac	agc	cgc	ttc	cgg	aat	agc	4606
Ala	Asp	Gly	Gln	Tyr	Asp	Pro	Tyr	Thr	Asp	Ser	Arg	Phe	Arg	Asn	Ser	
tcc	atg	tcc	ctg	gat	gag	aag	agc	aga	acc	atg	agc	cgt	tca	ggc	tca	4654
Ser	Met	Ser	Leu	Asp	Glu	Lys	Ser	Arg	Thr	Met	Ser	Arg	Ser	Gly	Ser	
ttc	cgg	gat	ggg	ttt	gaa	gaa	gtt	cat	gga	tcc	tca	ctc	tcc	ttg	gtt	4702
Phe	Arg	Asp	Gly	Phe	Glu	Glu	Val	His	Gly	Ser	Ser	Leu	Ser	Leu	Val	
tcc	agc	aca	tcg	tca	gtt	tat	tct	aca	cca	gaa	gaa	aaa	tgc	cag	tca	4750
Ser	Ser	Thr	Ser	Ser	Val	Tyr	Ser	Thr	Pro	Glu	Glu	Lys	Cys	Gln	Ser	
gag	att	cgc	aag	ctg	cgg	cgg	gaa	ctg	gat	gcc	tcc	cag	gag	aaa	gtt	4798
Glu	Ile	Arg	Lys	Leu	Arg	Arg	Glu	Leu	Asp	Ala	Ser	Gln	Glu	Lys	Val	
tca	gct	ttg	acc	acc	cag	ctg	aca	gca	aat	gct	cac	ctt	gtg	gct	gcc	4846
Ser	Ala	Leu	Thr	Thr	Gln	Leu	Thr	Ala	Asn	Ala	His	Leu	Val	Ala	Ala	
ttt	gaa	cag	agt	ctt	ggg	aac	atg	aca	atc	agg	ctc	cag	agt	ctg	acc	4894
Phe	Glu	Gln	Ser	Leu	Gly	Asn	Met	Thr	Ile	Arg	Leu	Gln	Ser	Leu	Thr	
atg	aca	gct	gag	cag	aag	gat	tca	gaa	ctg	aat	gag	tta	aga	aaa	acc	4942
Met	Thr	Ala	Glu	Gln	Lys	Asp	Ser	Glu	Leu	Asn	Glu	Leu	Arg	Lys	Thr	
att	gag	ctg	cta	aag	aaa	cag	aac	gca	gct	gcc	cag	gct	gcc	att	aat	4990
Ile	Glu	Leu	Leu	Lys	Lys	Gln	Asn	Ala	Ala	Ala	Gln	Ala	Ala	Ile	Asn	
gga	gta	att	aac	aca	cct	gag	ctc	aac	tgc	aaa	gga	aac	ggc	act	gcc	5038
Gly	Val	Ile	Asn	Thr	Pro	Glu	Leu	Asn	Cys	Lys	Gly	Asn	Gly	Thr	Ala	
cag	tct	gca	gac	ctc	cgc	atc	cgc	agg	cag	cac	tcc	tca	gac	agc	gtc	5086
Gln	Ser	Ala	Asp	Leu	Arg	Ile	Arg	Arg	Gln	His	Ser	Ser	Asp	Ser	Val	
tcc	agc	atc	aac	agt	gcc	acc	agc	cac	tcc	agt	gtg	ggc	agc	aac	ata	5134
Ser	Ser	Ile	Asn	Ser	Ala	Thr	Ser	His	Ser	Ser	Val	Gly	Ser	Asn	Ile	
gag	agt	gac	tca	aag	aag	aag	aag	agg	aag	aac	tgg	gtc	aat	gag	tta	5182
Glu	Ser	Asp	Ser	Lys	Lys	Lys	Lys	Arg	Lys	Asn	Trp	Val	Asn	Glu	Leu	
cgc	agc	tcc	ttc	aag	caa	gct	ttc	ggg	aag	aag	aag	tcc	cca	aaa	tct	5230
Arg	Ser	Ser	Phe	Lys	Gln	Ala	Phe	Gly	Lys	Lys	Lys	Ser	Pro	Lys	Ser	

gcg tcc tct cat tca gat att gag gag atg acg gat tct tct ttg cct Ala Ser Ser His Ser Asp Ile Glu Glu Met Thr Asp Ser Ser Leu Pro	5278
tcc tca cca aag tta cca cac aat ggg tcc aca ggt tcc acc cca ctg Ser Ser Pro Lys Leu Pro His Asn Gly Ser Thr Gly Ser Thr Pro Leu	5326
ctg agg aat tct cac tcc aac tct cta att tca gaa tgc atg gat agt Leu Arg Asn Ser His Ser Asn Ser Leu Ile Ser Glu Cys Met Asp Ser	5374
gaa gct gag acc gtc atg cag ctc cga aat gag tta aga gac aag gag Glu Ala Glu Thr Val Met Gln Leu Arg Asn Glu Leu Arg Asp Lys Glu	5422
atg aag ctg aca gat atc cgc tta gaa gct ctc agt tct gcc cac cag Met Lys Leu Thr Asp Ile Arg Leu Glu Ala Leu Ser Ser Ala His Gln	5470
ctg gac cag ctc cgg gag gcc atg aac agg atg cag agt gaa ata gag Leu Asp Gln Leu Arg Glu Ala Met Asn Arg Met Gln Ser Glu Ile Glu	5518
aag ctg aaa gct gag aat gat cgg ctg aag tca gag tct caa ggc agt Lys Leu Lys Ala Glu Asn Asp Arg Leu Lys Ser Glu Ser Gln Gly Ser	5566
ggc tgc agc cgg gct cct tcc caa gtg tcc atc tct gcc tcc ccg agg Gly Cys Ser Arg Ala Pro Ser Gln Val Ser Ile Ser Ala Ser Pro Arg	5614
cag tcc atg ggc ctc tcc cag cac agc ttg aac ctc act gag tca acc Gln Ser Met Gly Leu Ser Gln His Ser Leu Asn Leu Thr Glu Ser Thr	5662
agc ctg gac atg ttg ctg gat gac act ggt gaa tgc tcg gct cgg aag Ser Leu Asp Met Leu Leu Asp Asp Thr Gly Glu Cys Ser Ala Arg Lys	5710
gaa gga ggc agg cat gtt aag ata gtt gtc agc ttt cag gag gaa atg Glu Gly Gly Arg His Val Lys Ile Val Val Ser Phe Gln Glu Glu Met	5758
aag tgg aag gag gat tcc aga cca cat ctc ttt ctt att ggc tgc att Lys Trp Lys Glu Asp Ser Arg Pro His Leu Phe Leu Ile Gly Cys Ile	5806
gga gtt agt ggc aag acg aag tgg gat gtg ctc gat ggg gtg gtt aga Gly Val Ser Gly Lys Thr Lys Trp Asp Val Leu Asp Gly Val Val Arg	5854
cgg ctg ttc aaa gaa tac atc att cat gtc gac cca gtg agt cag cta Arg Leu Phe Lys Glu Tyr Ile Ile His Val Asp Pro Val Ser Gln Leu	5902
ggg ctg aat tca gac agc gtt ctt ggc tac agc att gga gaa atc aag Gly Leu Asn Ser Asp Ser Val Leu Gly Tyr Ser Ile Gly Glu Ile Lys	5950

cgc	agc	aac	act	tcc	gaa	aca	ccg	gag	ctg	ctt	cct	tgt	ggc	tat	ctg	5998
Arg	Ser	Asn	Thr	Ser	Glu	Thr	Pro	Glu	Leu	Leu	Pro	Cys	Gly	Tyr	Leu	
gtt	gga	gag	aac	acg	acc	atc	tca	gtg	act	gtg	aaa	ggg	ctc	gca	gaa	6046
Val	Gly	Glu	Asn	Thr	Thr	Ile	Ser	Val	Thr	Val	Lys	Gly	Leu	Ala	Glu	
aac	agc	ctg	gac	tca	ctg	gtg	ttt	gag	tcc	ttg	att	ccc	aag	ccc	atc	6094
Asn	Ser	Leu	Asp	Ser	Leu	Val	Phe	Glu	Ser	Leu	Ile	Pro	Lys	Pro	Ile	
ctg	cag	cgc	tac	gtc	tcc	ctc	ctg	ata	gag	cac	cgt	cgg	atc	att	ctc	6142
Leu	Gln	Arg	Tyr	Val	Ser	Leu	Leu	Ile	Glu	His	Arg	Arg	Ile	Ile	Leu	
tct	ggc	ccc	agc	ggc	act	ggg	aaa	acc	tac	ctg	gcc	aac	cgg	ctg	tct	6190
Ser	Gly	Pro	Ser	Gly	Thr	Gly	Lys	Thr	Tyr	Leu	Ala	Asn	Arg	Leu	Ser	
gag	tat	ata	gtg	ctt	cga	gag	gga	cgg	gag	ttg	aca	gac	ggg	gtt	atc	6238
Glu	Tyr	Ile	Val	Leu	Arg	Glu	Gly	Arg	Glu	Leu	Thr	Asp	Gly	Val	Ile	
gcc	acc	ttt	aac	gtg	gac	cat	aag	tcc	agc	aag	gaa	ttg	cgc	cag	tac	6286
Ala	Thr	Phe	Asn	Val	Asp	His	Lys	Ser	Ser	Lys	Glu	Leu	Arg	Gln	Tyr	
ctg	tcc	aac	ctt	gct	gac	cag	tgc	aac	agt	gag	aac	aat	gct	gtg	gac	6334
Leu	Ser	Asn	Leu	Ala	Asp	Gln	Cys	Asn	Ser	Glu	Asn	Asn	Ala	Val	Asp	
atg	ccc	ctc	gtc	atc	atc	ctg	gac	aac	cta	cac	cac	gtg	agc	tct	ctg	6382
Met	Pro	Leu	Val	Ile	Ile	Leu	Asp	Asn	Leu	His	His	Val	Ser	Ser	Leu	
ggc	gag	atc	ttc	aat	ggg	ctg	ctc	aac	tgc	aag	tac	cac	aaa	tgc	cct	6430
Gly	Glu	Ile	Phe	Asn	Gly	Leu	Leu	Asn	Cys	Lys	Tyr	His	Lys	Cys	Pro	
tac	ata	att	ggc	aca	atg	aac	cag	gct	acc	tct	tcg	act	ccc	aac	ctg	6478
Tyr	Ile	Ile	Gly	Thr	Met	Asn	Gln	Ala	Thr	Ser	Ser	Thr	Pro	Asn	Leu	
cag	ctt	cac	cat	aac	ttc	aga	tgg	gtg	ctt	tgt	gcc	aac	cac	acg	gag	6526
Gln	Leu	His	His	Asn	Phe	Arg	Trp	Val	Leu	Cys	Ala	Asn	His	Thr	Glu	
cct	gtg	aag	ggg	ttc	ctt	ggc	cga	ttc	ctg	agg	agg	aag	ctc	atg	gaa	6574
Pro	Val	Lys	Gly	Phe	Leu	Gly	Arg	Phe	Leu	Arg	Arg	Lys	Leu	Met	Glu	
aca	gag	atc	agt	ggg	cgg	gtg	cgc	aat	atg	gag	ctg	gta	aaa	atc	att	6622
Thr	Glu	Ile	Ser	Gly	Arg	Val	Arg	Asn	Met	Glu	Leu	Val	Lys	Ile	Ile	

gac tgg att ccc aag gtc tgg cat cac ctc aac cgc ttc ctg gag gct Asp Trp Ile Pro Lys Val Trp His His Leu Asn Arg Phe Leu Glu Ala	6670
cac agt tcc tcg gac gtc acc atc ggc ccc cgg ctc ttc ctg tca tgc His Ser Ser Ser Asp Val Thr Ile Gly Pro Arg Leu Phe Leu Ser Cys	6718
ccc atc gat gtg gac ggc tcg aga gtg tgg ttc acc gac ttg tgg aac Pro Ile Asp Val Asp Gly Ser Arg Val Trp Phe Thr Asp Leu Trp Asn	6766
tat tcc att atc ccc tat ctc ctg gaa gcc gtc aga gaa gga ctc cag Tyr Ser Ile Ile Pro Tyr Leu Leu Glu Ala Val Arg Glu Gly Leu Gln	6814
ctc tat gga agg cgc gcc ccc tgg gag gat cct gcc aag tgg gtg atg Leu Tyr Gly Arg Arg Ala Pro Trp Glu Asp Pro Ala Lys Trp Val Met	6862
gac aca tat cca tgg gca gcc agc cca caa cag cac gag tgg cct ccc Asp Thr Tyr Pro Trp Ala Ala Ser Pro Gln Gln His Glu Trp Pro Pro	6910
ctg ctg cag tta cgg cct gag gat gtc ggc ttc gac ggc tac tcc atg Leu Leu Gln Leu Arg Pro Glu Asp Val Gly Phe Asp Gly Tyr Ser Met	6958
cct cgg gag gga tcg aca agc aag cag atg ccc ccc agt gat gct gaa Pro Arg Glu Gly Ser Thr Ser Lys Gln Met Pro Pro Ser Asp Ala Glu	7006
ggt gac ccg ctg atg aac atg ctg atg agg ctg cag gag gca gcc aac Gly Asp Pro Leu Met Asn Met Leu Met Arg Leu Gln Glu Ala Ala Asn	7054
tac tcc agc ccc cag agc tat gac agc gac tcc aac agc aac agc cat Tyr Ser Ser Pro Gln Ser Tyr Asp Ser Asp Ser Asn Ser Asn Ser His	7102
cac gat gac atc ttg gac tcc tct ttg gag tcc act ctg tgacaggggc His Asp Asp Ile Leu Asp Ser Ser Leu Glu Ser Thr Leu	7151
ccggagccca gcgcctcct cttctcctca ccgcattcca cctgcatccc ccacatcacc	7211
ctgaagatga cttcctgagc cagccccag ccacagcctt agagctgcgg gaacaccgag	7271
accccccgtc cttcagcctc gacctgggtg caggcatccc gggccagctg cctgcggacc	7331
gcttccttcc acagcgagaa ctgcactacc ttctgttgta ctttaattat tgttttgcc	7391
tgttgctgtg acctccctaa gacactgaag atacttctcg ggaaaggatc atcgccgttg	7451
aaatgaaaag agagacagag agagaaaaaa aaaagagaac ccacatgaag ctctgaaacc	7511

aaacagcatc ctgccatgag cttcccagag acagaagaga ctggagcaaa gtcggaaaca 7571  
 cagagaagca cggttcccc tcagcacaga cctccagac tgggtctcag agccgtgcca 7631  
 cccaccctcc cacacagccg gccacagggg gaactggtgc taaccagggt gcttgctttg 7691  
 gtcacgttca acgcactaca gagctacgac acaggggaac cttaggagca aataaacctg 7751  
 gctttcatgt tttttaaaaa aaaaaaaaaa aa 7783

<210> 27  
 <211> 2380  
 <212> PRT  
 <213> Homo sapiens  
 <400> 27

Asn	Gln	Pro	Glu	Arg	Leu	Asn	Ser	Gln	Val	Leu	Gln	Gly	Leu	Gln	Glu	
1				5					10					15		
Pro	Ala	Gly	Glu	Gly	Leu	Pro	Leu	Arg	Lys	Ser	Gly	Ser	Val	Glu	Asn	
			20					25					30			
Gly	Phe	Asp	Thr	Gln	Ile	Tyr	Thr	Asp	Trp	Ala	Asn	His	Tyr	Leu	Ala	
		35					40					45				
Lys	Ser	Gly	His	Lys	Arg	Leu	Ile	Arg	Asp	Leu	Gln	Gln	Asp	Val	Thr	
	50					55					60					
Asp	Gly	Val	Leu	Leu	Ala	Gln	Ile	Ile	Gln	Val	Val	Ala	Asn	Glu	Lys	
65					70				75					80		
Ile	Glu	Asp	Ile	Asn	Gly	Cys	Pro	Lys	Asn	Arg	Ser	Gln	Met	Ile	Glu	
				85					90					95		
Asn	Ile	Asp	Ala	Cys	Leu	Asn	Phe	Leu	Ala	Ala	Lys	Gly	Ile	Asn	Ile	
			100					105					110			
Gln	Gly	Leu	Ser	Ala	Glu	Glu	Ile	Arg	Asn	Gly	Asn	Leu	Lys	Ala	Ile	
		115					120					125				
Leu	Gly	Leu	Phe	Phe	Ser	Leu	Ser	Arg	Tyr	Lys	Gln	Gln	Gln	Gln	Gln	
	130					135					140					
Pro	Gln	Lys	Gln	His	Leu	Ser	Ser	Pro	Leu	Pro	Pro	Ala	Val	Ser	Gln	
145					150					155				160		
Val	Ala	Gly	Ala	Pro	Ser	Gln	Cys	Gln	Ala	Gly	Thr	Pro	Gln	Gln	Gln	
				165				170					175			
Val	Pro	Val	Thr	Pro	Gln	Ala	Pro	Cys	Gln	Pro	His	Gln	Pro	Ala	Pro	
			180					185				190				
His	Gln	Gln	Ser	Lys	Ala	Gln	Ala	Glu	Met	Gln	Ser	Arg	Leu	Pro	Gly	



195					200					205					
Pro	Thr	Ala	Arg	Val	Ser	Ala	Ala	Gly	Ser	Glu	Ala	Lys	Thr	Arg	Gly
210						215					220				
Gly	Ser	Thr	Thr	Ala	Asn	Asn	Arg	Arg	Ser	Gln	Ser	Phe	Asn	Asn	Tyr
225					230					235					240
Asp	Lys	Ser	Lys	Pro	Val	Thr	Ser	Pro	Pro	Pro	Pro	Pro	Ser	Ser	His
				245					250					255	
Glu	Lys	Glu	Pro	Leu	Ala	Ser	Ser	Ala	Ser	Ser	His	Pro	Gly	Met	Ser
			260					265					270		
Asp	Asn	Ala	Pro	Ala	Ser	Leu	Glu	Ser	Gly	Ser	Ser	Ser	Thr	Pro	Thr
		275					280						285		
Asn	Cys	Ser	Thr	Ser	Ser	Ala	Ile	Pro	Gln	Pro	Gly	Ala	Ala	Thr	Lys
	290					295					300				
Pro	Trp	Arg	Ser	Lys	Ser	Leu	Ser	Val	Lys	His	Ser	Ala	Thr	Val	Ser
305					310					315					320
Met	Leu	Ser	Val	Lys	Pro	Pro	Gly	Pro	Glu	Ala	Pro	Arg	Pro	Thr	Pro
				325					330					335	
Glu	Ala	Met	Lys	Pro	Ala	Pro	Asn	Asn	Gln	Lys	Ser	Met	Leu	Glu	Lys
			340					345					350		
Leu	Lys	Leu	Phe	Asn	Ser	Lys	Gly	Gly	Ser	Lys	Ala	Gly	Glu	Gly	Pro
		355					360					365			
Gly	Ser	Arg	Asp	Thr	Ser	Cys	Glu	Arg	Leu	Glu	Thr	Leu	Pro	Ser	Phe
	370					375					380				
Glu	Glu	Ser	Glu	Glu	Leu	Glu	Ala	Ala	Ser	Arg	Met	Leu	Thr	Thr	Val
385					390					395					400
Gly	Pro	Ala	Ser	Ser	Ser	Pro	Lys	Ile	Ala	Leu	Lys	Gly	Ile	Ala	Gln
				405					410					415	
Arg	Thr	Phe	Ser	Arg	Ala	Leu	Thr	Asn	Lys	Lys	Ser	Ser	Leu	Lys	Gly
			420					425					430		
Asn	Glu	Lys	Glu	Lys	Glu	Lys	Gln	Gln	Arg	Glu	Lys	Asp	Lys	Glu	Lys
		435					440					445			
Ser	Lys	Asp	Leu	Ala	Lys	Arg	Ala	Ser	Val	Thr	Glu	Arg	Leu	Asp	Leu
	450					455					460				
Lys	Glu	Glu	Pro	Lys	Glu	Asp	Pro	Ser	Gly	Ala	Ala	Val	Pro	Glu	Met
465					470					475					480
Pro	Lys	Lys	Ser	Ser	Lys	Ile	Ala	Ser	Phe	Ile	Pro	Lys	Gly	Gly	Lys
				485					490					495	

Leu	Asn	Ser	Ala	Lys	Lys	Glu	Pro	Met	Ala	Pro	Ser	His	Ser	Gly	Ile
			500												
Pro	Lys	Pro	Gly	Met	Lys	Ser	Met	Pro	Gly	Lys	Ser	Pro	Ser	Ala	Pro
			515												
Ala	Pro	Ser	Lys	Glu	Gly	Glu	Arg	Ser	Arg	Ser	Gly	Lys	Leu	Ser	Ser
			530												
Gly	Leu	Pro	Gln	Gln	Lys	Pro	Gln	Leu	Asp	Gly	Arg	His	Ser	Ser	Ser
			545												
Ser	Ser	Ser	Leu	Ala	Ser	Ser	Glu	Gly	Lys	Gly	Pro	Gly	Gly	Thr	Thr
			565												
Leu	Asn	His	Ser	Ile	Ser	Ser	Gln	Thr	Val	Ser	Gly	Ser	Val	Gly	Thr
			580												
Thr	Gln	Thr	Thr	Gly	Ser	Asn	Thr	Val	Ser	Val	Gln	Leu	Pro	Gln	Pro
			595												
Gln	Gln	Gln	Tyr	Asn	His	Pro	Asn	Thr	Ala	Thr	Val	Ala	Pro	Phe	Leu
			610												
Tyr	Arg	Ser	Gln	Thr	Asp	Thr	Glu	Gly	Asn	Val	Thr	Ala	Glu	Ser	Ser
			625												
Ser	Thr	Gly	Val	Ser	Val	Glu	Pro	Ser	His	Phe	Thr	Lys	Thr	Gly	Gln
			645												
Pro	Ala	Leu	Glu	Glu	Leu	Thr	Gly	Glu	Asp	Pro	Glu	Ala	Arg	Arg	Leu
			660												
Arg	Thr	Val	Lys	Asn	Ile	Ala	Asp	Leu	Arg	Gln	Asn	Leu	Glu	Glu	Thr
			675												
Met	Ser	Ser	Leu	Arg	Gly	Thr	Gln	Val	Thr	His	Ser	Thr	Leu	Glu	Thr
			690												
Thr	Phe	Asp	Thr	Asn	Val	Thr	Thr	Glu	Met	Ser	Gly	Arg	Ser	Ile	Leu
			705												
Ser	Leu	Thr	Gly	Arg	Pro	Thr	Pro	Leu	Ser	Trp	Arg	Leu	Gly	Gln	Ser
			725												
Ser	Pro	Arg	Leu	Gln	Ala	Gly	Asp	Ala	Pro	Ser	Met	Gly	Asn	Gly	Tyr
			740												
Pro	Pro	Arg	Ala	Asn	Ala	Ser	Arg	Phe	Ile	Asn	Thr	Glu	Ser	Gly	Arg
			755												
Tyr	Val	Tyr	Ser	Ala	Pro	Leu	Arg	Arg	Gln	Leu	Ala	Ser	Arg	Gly	Ser
			770												
Ser	Val	Cys	His	Val	Asp	Val	Ser	Asp	Lys	Ala	Gly	Asp	Glu	Met	Asp
			785												

Leu	Glu	Gly	Ile	Ser	Met	Asp	Ala	Pro	Gly	Tyr	Met	Ser	Asp	Gly	Asp
				805					810					815	
Val	Leu	Ser	Lys	Asn	Ile	Arg	Thr	Asp	Asp	Ile	Thr	Ser	Gly	Tyr	Met
			820					825					830		
Thr	Asp	Gly	Gly	Leu	Gly	Leu	Tyr	Thr	Arg	Arg	Leu	Asn	Arg	Leu	Pro
		835					840					845			
Asp	Gly	Met	Ala	Val	Val	Arg	Glu	Thr	Leu	Gln	Arg	Asn	Thr	Ser	Leu
	850					855					860				
Gly	Leu	Gly	Asp	Ala	Asp	Ser	Trp	Asp	Asp	Ser	Ser	Ser	Val	Ser	Ser
865					870					875					880
Gly	Ile	Ser	Asp	Thr	Ile	Asp	Asn	Leu	Ser	Thr	Asp	Asp	Ile	Asn	Thr
				885					890					895	
Ser	Ser	Ser	Ile	Ser	Ser	Tyr	Ala	Asn	Thr	Pro	Ala	Ser	Ser	Arg	Lys
			900					905					910		
Asn	Leu	Asp	Val	Gln	Thr	Asp	Ala	Glu	Lys	His	Ser	Gln	Val	Glu	Arg
		915					920					925			
Asn	Ser	Leu	Trp	Ser	Gly	Asp	Asp	Val	Lys	Lys	Ser	Asp	Gly	Gly	Ser
	930					935					940				
Asp	Ser	Gly	Ile	Lys	Met	Glu	Pro	Gly	Ser	Lys	Trp	Arg	Arg	Asn	Pro
945				950						955					960
Ser	Asp	Val	Ser	Asp	Glu	Ser	Asp	Lys	Ser	Thr	Ser	Gly	Lys	Lys	Asn
				965					970					975	
Pro	Val	Ile	Ser	Gln	Thr	Gly	Ser	Trp	Arg	Arg	Gly	Met	Thr	Ala	Gln
			980					985					990		
Val	Gly	Ile	Thr	Met	Pro	Arg	Thr	Lys	Ala	Ser	Ala	Pro	Ala	Gly	Ala
	995						1000					1005			
Leu	Lys	Thr	Pro	Gly	Thr	Gly	Lys	Thr	Asp	Asp	Ala	Lys	Val	Ser	Glu
1010						1015					1020				
Lys	Gly	Arg	Leu	Ser	Pro	Lys	Ala	Ser	Gln	Val	Lys	Arg	Ser	Pro	Ser
1025					1030					1035					1040
Asp	Ala	Gly	Arg	Ser	Ser	Gly	Asp	Glu	Ser	Lys	Lys	Pro	Leu	Pro	Ser
				1045					1050					1055	
Ser	Ser	Arg	Thr	Pro	Thr	Ala	Asn	Ala	Asn	Ser	Phe	Gly	Phe	Lys	Lys
			1060					1065					1070		
Gln	Ser	Gly	Ser	Ala	Thr	Gly	Leu	Ala	Met	Ile	Thr	Ala	Ser	Gly	Val
	1075						1080					1085			
Thr	Val	Thr	Ser	Arg	Ser	Ala	Thr	Leu	Gly	Lys	Ile	Pro	Lys	Ser	Ser

1090					1095					1100					
Ala	Leu	Val	Ser	Arg	Ser	Ala	Gly	Arg	Lys	Ser	Ser	Met	Asp	Gly	Ala
1105					1110					1115					1120
Gln	Asn	Gln	Asp	Asp	Gly	Tyr	Leu	Ala	Leu	Ser	Ser	Arg	Thr	Asn	Leu
				1125					1130					1135	
Gln	Tyr	Arg	Ser	Leu	Pro	Arg	Pro	Ser	Lys	Ser	Asn	Ser	Arg	Asn	Gly
			1140					1145					1150		
Ala	Gly	Asn	Arg	Ser	Ser	Thr	Ser	Ser	Ile	Asp	Ser	Asn	Ile	Ser	Ser
		1155					1160					1165			
Lys	Ser	Ala	Gly	Leu	Pro	Val	Pro	Lys	Leu	Arg	Glu	Pro	Ser	Lys	Thr
	1170					1175					1180				
Ala	Leu	Gly	Ser	Ser	Leu	Pro	Gly	Leu	Val	Asn	Gln	Thr	Asp	Lys	Glu
1185					1190					1195					1200
Lys	Gly	Ile	Ser	Ser	Asp	Asn	Glu	Ser	Val	Ala	Ser	Cys	Asn	Ser	Val
				1205					1210					1215	
Lys	Val	Asn	Pro	Ala	Ala	Gln	Pro	Val	Ser	Ser	Pro	Ala	Gln	Thr	Ser
			1220					1225					1230		
Leu	Gln	Pro	Gly	Ala	Lys	Tyr	Pro	Asp	Val	Ala	Ser	Pro	Thr	Leu	Arg
		1235					1240					1245			
Arg	Leu	Phe	Gly	Gly	Lys	Pro	Thr	Lys	Gln	Val	Pro	Ile	Ala	Thr	Ala
	1250					1255					1260				
Glu	Asn	Met	Lys	Asn	Ser	Val	Val	Ile	Ser	Asn	Pro	His	Ala	Thr	Met
1265					1270					1275					1280
Thr	Gln	Gln	Gly	Asn	Leu	Asp	Ser	Pro	Ser	Gly	Ser	Gly	Val	Leu	Ser
				1285					1290					1295	
Ser	Gly	Ser	Ser	Ser	Pro	Leu	Tyr	Ser	Lys	Asn	Val	Asp	Leu	Asn	Gln
			1300					1305					1310		
Ser	Pro	Leu	Ala	Ser	Ser	Pro	Ser	Ser	Ala	His	Ser	Ala	Pro	Ser	Asn
		1315					1320					1325			
Ser	Leu	Thr	Trp	Gly	Thr	Asn	Ala	Ser	Ser	Ser	Ser	Ala	Val	Ser	Lys
	1330					1335					1340				
Asp	Gly	Leu	Gly	Phe	Gln	Ser	Val	Ser	Ser	Leu	His	Thr	Ser	Cys	Glu
1345					1350					1355					1360
Ser	Ile	Asp	Ile	Ser	Leu	Ser	Ser	Gly	Gly	Val	Pro	Ser	His	Asn	Ser
				1365				1370						1375	
Ser	Thr	Gly	Leu	Ile	Ala	Ser	Ser	Lys	Asp	Asp	Ser	Leu	Thr	Pro	Phe
			1380					1385					1390		
Val	Arg	Thr	Asn	Ser	Val	Lys	Thr	Thr	Leu	Ser	Glu	Ser	Pro	Leu	Ser

1395	1400	1405
Ser Pro Ala Ala Ser Pro Lys Phe Cys Arg Ser Thr Leu Pro Arg Lys 1410 1415 1420		
Gln Asp Ser Asp Pro His Leu Asp Arg Asn Thr Leu Pro Lys Lys Gly 1425 1430 1435 1440		
Leu Arg Tyr Thr Pro Thr Ser Gln Leu Arg Thr Gln Glu Asp Ala Lys 1445 1450 1455		
Glu Trp Leu Arg Ser His Ser Ala Gly Gly Leu Gln Asp Thr Ala Ala 1460 1465 1470		
Asn Ser Pro Phe Ser Ser Gly Ser Ser Val Thr Ser Pro Ser Gly Thr 1475 1480 1485		
Arg Phe Asn Phe Ser Gln Leu Ala Ser Pro Thr Thr Val Thr Gln Met 1490 1495 1500		
Ser Leu Ser Asn Pro Thr Met Leu Arg Thr His Ser Leu Ser Asn Ala 1505 1510 1515 1520		
Asp Gly Gln Tyr Asp Pro Tyr Thr Asp Ser Arg Phe Arg Asn Ser Ser 1525 1530 1535		
Met Ser Leu Asp Glu Lys Ser Arg Thr Met Ser Arg Ser Gly Ser Phe 1540 1545 1550		
Arg Asp Gly Phe Glu Glu Val His Gly Ser Ser Leu Ser Leu Val Ser 1555 1560 1565		
Ser Thr Ser Ser Val Tyr Ser Thr Pro Glu Glu Lys Cys Gln Ser Glu 1570 1575 1580		
Ile Arg Lys Leu Arg Arg Glu Leu Asp Ala Ser Gln Glu Lys Val Ser 1585 1590 1595 1600		
Ala Leu Thr Thr Gln Leu Thr Ala Asn Ala His Leu Val Ala Ala Phe 1605 1610 1615		
Glu Gln Ser Leu Gly Asn Met Thr Ile Arg Leu Gln Ser Leu Thr Met 1620 1625 1630		
Thr Ala Glu Gln Lys Asp Ser Glu Leu Asn Glu Leu Arg Lys Thr Ile 1635 1640 1645		
Glu Leu Leu Lys Lys Gln Asn Ala Ala Ala Gln Ala Ala Ile Asn Gly 1650 1655 1660		
Val Ile Asn Thr Pro Glu Leu Asn Cys Lys Gly Asn Gly Thr Ala Gln 1665 1670 1675 1680		
Ser Ala Asp Leu Arg Ile Arg Arg Gln His Ser Ser Asp Ser Val Ser 1685 1690 1695		

Ser Ile Asn Ser Ala Thr Ser His Ser Ser Val Gly Ser Asn Ile Glu  
 1700 1705 1710  
 Ser Asp Ser Lys Lys Lys Lys Arg Lys Asn Trp Val Asn Glu Leu Arg  
 1715 1720 1725  
 Ser Ser Phe Lys Gln Ala Phe Gly Lys Lys Lys Ser Pro Lys Ser Ala  
 1730 1735 1740  
 Ser Ser His Ser Asp Ile Glu Glu Met Thr Asp Ser Ser Leu Pro Ser  
 1745 1750 1755 1760  
 Ser Pro Lys Leu Pro His Asn Gly Ser Thr Gly Ser Thr Pro Leu Leu  
 1765 1770 1775  
 Arg Asn Ser His Ser Asn Ser Leu Ile Ser Glu Cys Met Asp Ser Glu  
 1780 1785 1790  
 Ala Glu Thr Val Met Gln Leu Arg Asn Glu Leu Arg Asp Lys Glu Met  
 1795 1800 1805  
 Lys Leu Thr Asp Ile Arg Leu Glu Ala Leu Ser Ser Ala His Gln Leu  
 1810 1815 1820  
 Asp Gln Leu Arg Glu Ala Met Asn Arg Met Gln Ser Glu Ile Glu Lys  
 1825 1830 1835 1840  
 Leu Lys Ala Glu Asn Asp Arg Leu Lys Ser Glu Ser Gln Gly Ser Gly  
 1845 1850 1855  
 Cys Ser Arg Ala Pro Ser Gln Val Ser Ile Ser Ala Ser Pro Arg Gln  
 1860 1865 1870  
 Ser Met Gly Leu Ser Gln His Ser Leu Asn Leu Thr Glu Ser Thr Ser  
 1875 1880 1885  
 Leu Asp Met Leu Leu Asp Asp Thr Gly Glu Cys Ser Ala Arg Lys Glu  
 1890 1895 1900  
 Gly Gly Arg His Val Lys Ile Val Val Ser Phe Gln Glu Glu Met Lys  
 1905 1910 1915 1920  
 Trp Lys Glu Asp Ser Arg Pro His Leu Phe Leu Ile Gly Cys Ile Gly  
 1925 1930 1935  
 Val Ser Gly Lys Thr Lys Trp Asp Val Leu Asp Gly Val Val Arg Arg  
 1940 1945 1950  
 Leu Phe Lys Glu Tyr Ile Ile His Val Asp Pro Val Ser Gln Leu Gly  
 1955 1960 1965  
 Leu Asn Ser Asp Ser Val Leu Gly Tyr Ser Ile Gly Glu Ile Lys Arg  
 1970 1975 1980  
 Ser Asn Thr Ser Glu Thr Pro Glu Leu Leu Pro Cys Gly Tyr Leu Val  
 1985 1990 1995 2000

Gly Glu Asn Thr Thr Ile Ser Val Thr Val Lys Gly Leu Ala Glu Asn  
 2005 2010 2015  
 Ser Leu Asp Ser Leu Val Phe Glu Ser Leu Ile Pro Lys Pro Ile Leu  
 2020 2025 2030  
 Gln Arg Tyr Val Ser Leu Leu Ile Glu His Arg Arg Ile Ile Leu Ser  
 2035 2040 2045  
 Gly Pro Ser Gly Thr Gly Lys Thr Tyr Leu Ala Asn Arg Leu Ser Glu  
 2050 2055 2060  
 Tyr Ile Val Leu Arg Glu Gly Arg Glu Leu Thr Asp Gly Val Ile Ala  
 2065 2070 2075 2080  
 Thr Phe Asn Val Asp His Lys Ser Ser Lys Glu Leu Arg Gln Tyr Leu  
 2085 2090 2095  
 Ser Asn Leu Ala Asp Gln Cys Asn Ser Glu Asn Asn Ala Val Asp Met  
 2100 2105 2110  
 Pro Leu Val Ile Ile Leu Asp Asn Leu His His Val Ser Ser Leu Gly  
 2115 2120 2125  
 Glu Ile Phe Asn Gly Leu Leu Asn Cys Lys Tyr His Lys Cys Pro Tyr  
 2130 2135 2140  
 Ile Ile Gly Thr Met Asn Gln Ala Thr Ser Ser Thr Pro Asn Leu Gln  
 2145 2150 2155 2160  
 Leu His His Asn Phe Arg Trp Val Leu Cys Ala Asn His Thr Glu Pro  
 2165 2170 2175  
 Val Lys Gly Phe Leu Gly Arg Phe Leu Arg Arg Lys Leu Met Glu Thr  
 2180 2185 2190  
 Glu Ile Ser Gly Arg Val Arg Asn Met Glu Leu Val Lys Ile Ile Asp  
 2195 2200 2205  
 Trp Ile Pro Lys Val Trp His His Leu Asn Arg Phe Leu Glu Ala His  
 2210 2215 2220  
 Ser Ser Ser Asp Val Thr Ile Gly Pro Arg Leu Phe Leu Ser Cys Pro  
 2225 2230 2235 2240  
 Ile Asp Val Asp Gly Ser Arg Val Trp Phe Thr Asp Leu Trp Asn Tyr  
 2245 2250 2255  
 Ser Ile Ile Pro Tyr Leu Leu Glu Ala Val Arg Glu Gly Leu Gln Leu  
 2260 2265 2270  
 Tyr Gly Arg Arg Ala Pro Trp Glu Asp Pro Ala Lys Trp Val Met Asp  
 2275 2280 2285  
 Thr Tyr Pro Trp Ala Ala Ser Pro Gln Gln His Glu Trp Pro Pro Leu

2290                      2295                      2300  
 Leu Gln Leu Arg Pro Glu Asp Val Gly Phe Asp Gly Tyr Ser Met Pro  
 2305                      2310                      2315                      2320  
 Arg Glu Gly Ser Thr Ser Lys Gln Met Pro Pro Ser Asp Ala Glu Gly  
                          2325                      2330                      2335  
 Asp Pro Leu Met Asn Met Leu Met Arg Leu Gln Glu Ala Ala Asn Tyr  
                          2340                      2345                      2350  
 Ser Ser Pro Gln Ser Tyr Asp Ser Asp Ser Asn Ser Asn Ser His His  
                          2355                      2360                      2365  
 Asp Asp Ile Leu Asp Ser Ser Leu Glu Ser Thr Leu  
                          2370                      2375                      2380

<210> 28  
 <211> 96  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (1)..(93)

<400> 28

agc agg gag agg gga ggg agt gtg ccg tct ctt ctg caa ggg cag tgc                      48  
 Ser Arg Glu Arg Gly Gly Ser Val Pro Ser Leu Leu Gln Gly Gln Cys  
 ccc agc ctc agc cac act tct gat ctg cag tcc aac aga cct ttc tag                      96  
 Pro Ser Leu Ser His Thr Ser Asp Leu Gln Ser Asn Arg Pro Phe \*

<210> 29  
 <211> 75  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (1)..(72)

<400> 29

cat gcc aaa gag aac ctg ggg gtg cca ggg ggt cct cag agc tca cac                      48  
 His Ala Lys Glu Asn Leu Gly Val Pro Gly Gly Pro Gln Ser Ser His  
 tgc act tgt ggc acc cac agc gag tag                      75  
 Cys Thr Cys Gly Thr His Ser Glu \*



<210> 30  
 <211> 489  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> CDS  
 <222> (1)..(489)

<400> 30

cca tcc gtg agc cga gga aac tgt aca cag atc tac aca gac tgg gcc	48
Pro Ser Val Ser Arg Gly Asn Cys Thr Gln Ile Tyr Thr Asp Trp Ala	
aat cat tac cta gcc aaa tcc ggc cac aag cgt ctc atc aag gat ctc	96
Asn His Tyr Leu Ala Lys Ser Gly His Lys Arg Leu Ile Lys Asp Leu	
cag caa gat gtg aca gat ggc gtc ctc ctg gcc cag att atc cag gtt	144
Gln Gln Asp Val Thr Asp Gly Val Leu Leu Ala Gln Ile Ile Gln Val	
gtg gca aat gaa aag att gaa gac atc aat ggc tgt ccg aag aac aga	192
Val Ala Asn Glu Lys Ile Glu Asp Ile Asn Gly Cys Pro Lys Asn Arg	
tcc caa atg att gaa aac ata gat gcc tgc ttg aat ttc ctg gca gct	240
Ser Gln Met Ile Glu Asn Ile Asp Ala Cys Leu Asn Phe Leu Ala Ala	
aag gga ata aac atc cag ggg ctg tct gca gaa gag atc agg aat gga	288
Lys Gly Ile Asn Ile Gln Gly Leu Ser Ala Glu Glu Ile Arg Asn Gly	
aac ctc aag gcc att cta ggc ctc ttc ttc agc ctc tcc cga tac aag	336
Asn Leu Lys Ala Ile Leu Gly Leu Phe Phe Ser Leu Ser Arg Tyr Lys	
cag cag cag cag cag ccc cag aag cag cac ctc tcc tca cct ctg ccg	384
Gln Gln Gln Gln Gln Pro Gln Lys Gln His Leu Ser Ser Pro Leu Pro	
ccc gcc gta tcc cag gtg gcc ggg gcc ccc tcc cag tgc cag gct ggc	432
Pro Ala Val Ser Gln Val Ala Gly Ala Pro Ser Gln Cys Gln Ala Gly	
acc cct cag cag cag gtg cca gtc act ccc caa gcc ccg tgc cag cct	480
Thr Pro Gln Gln Gln Val Pro Val Thr Pro Gln Ala Pro Cys Gln Pro	
cac cag cca	489
His Gln Pro	

<210> 31  
 <211> 31  
 <212> PRT  
 <213> Homo sapiens

<400> 31

Ser Arg Glu Arg Gly Gly Ser Val Pro Ser Leu Leu Gln Gly Gln Cys  
 1 5 10 15

Pro Ser Leu Ser His Thr Ser Asp Leu Gln Ser Asn Arg Pro Phe  
 20 25 30

<210> 32  
 <211> 24  
 <212> PRT  
 <213> Homo sapiens

<400> 32

His Ala Lys Glu Asn Leu Gly Val Pro Gly Gly Pro Gln Ser Ser His  
 5 10 15

Cys Thr Cys Gly Thr His Ser Glu  
 20

<210> 33  
 <211> 163  
 <212> PRT  
 <213> Homo sapiens

<400> 33

Pro Ser Val Ser Arg Gly Asn Cys Thr Gln Ile Tyr Thr Asp Trp Ala  
 5 10 15

Asn His Tyr Leu Ala Lys Ser Gly His Lys Arg Leu Ile Lys Asp Leu  
 20 25 30

Gln Gln Asp Val Thr Asp Gly Val Leu Leu Ala Gln Ile Ile Gln Val  
 35 40 45

Val Ala Asn Glu Lys Ile Glu Asp Ile Asn Gly Cys Pro Lys Asn Arg  
 50 55 60

Ser Gln Met Ile Glu Asn Ile Asp Ala Cys Leu Asn Phe Leu Ala Ala

65                      70                      75                      80  
 Lys Gly Ile Asn Ile Gln Gly Leu Ser Ala Glu Glu Ile Arg Asn Gly  
                                  85                      90                      95  
 Asn Leu Lys Ala Ile Leu Gly Leu Phe Phe Ser Leu Ser Arg Tyr Lys  
                                  100                      105                      110  
 Gln Gln Gln Gln Gln Pro Gln Lys Gln His Leu Ser Ser Pro Leu Pro  
                                  115                      120                      125  
 Pro Ala Val Ser Gln Val Ala Gly Ala Pro Ser Gln Cys Gln Ala Gly  
                                  130                      135                      140  
 Thr Pro Gln Gln Gln Val Pro Val Thr Pro Gln Ala Pro Cys Gln Pro  
 145                      150                      155                      160  
 His Gln Pro

<210> 34  
 <211> 479  
 <212> DNA  
 <213> mouse

<220>  
 <221> CDS  
 <222> (3)...(476)

<400> 34

at gag aag agc cga aca atg agt cgg tca ggc tcc ttc cgg gat ggg                      47  
     Glu Lys Ser Arg Thr Met Ser Arg Ser Gly Ser Phe Arg Asp Gly  
  
 ttt gag gaa gtt cat gga tcc tcc ctg tcc ttg gtt tcc agc aca tcc                      95  
 Phe Glu Glu Val His Gly Ser Ser Leu Ser Leu Val Ser Ser Thr Ser  
  
 tcc atc tac tcc acg cca gaa gaa aaa tgc cag tca gag att cga aag                      143  
 Ser Ile Tyr Ser Thr Pro Glu Glu Lys Cys Gln Ser Glu Ile Arg Lys  
  
 ctg agg cga gaa ctg gat gcc tcc cag gaa aag gtg tct gcg ctg act                      191  
 Leu Arg Arg Glu Leu Asp Ala Ser Gln Glu Lys Val Ser Ala Leu Thr  
  
 acc cag ctg act gca aat gct cac ctt gtg gca gcc ttc gag cag agt                      239  
 Thr Gln Leu Thr Ala Asn Ala His Leu Val Ala Ala Phe Glu Gln Ser  
  
 ctg gga aac atg acc atc agg cta cag agt tta act atg acc gct gag                      287  
 Leu Gly Asn Met Thr Ile Arg Leu Gln Ser Leu Thr Met Thr Ala Glu  
  
 cag aag gat tca gaa ctg aac gag tta aga aaa acc atc gag ctg ctg                      335

Gln Lys Asp Ser Glu Leu Asn Glu Leu Arg Lys Thr Ile Glu Leu Leu

aag aaa cag aat gca gct gcc cag gct gcc att aat gga gtg att aac 383  
Lys Lys Gln Asn Ala Ala Ala Gln Ala Ala Ile Asn Gly Val Ile Asn

acg cca gag ctc aac tgc aaa gga aat ggc agt gcc agg cta cag acc 431  
Thr Pro Glu Leu Asn Cys Lys Gly Asn Gly Ser Ala Arg Leu Gln Thr

tac gca tcc gca gca aca ctc ctc cga cag tgt ctc cag tat caa tag 479  
Tyr Ala Ser Ala Ala Thr Leu Leu Arg Gln Cys Leu Gln Tyr Gln \*

<210> 35  
<211> 22  
<212> DNA  
<213> mouse

<220>  
<221> CDS  
<222> (1)...(21)

<400> 35

cgc cac cag cca ctc aag tgt g 22  
Arg His Gln Pro Leu Lys Cys

<210> 36  
<211> 158  
<212> PRT  
<213> mouse

<400> 36

Glu Lys Ser Arg Thr Met Ser Arg Ser Gly Ser Phe Arg Asp Gly Phe  
1 5 10 15

Glu Glu Val His Gly Ser Ser Leu Ser Leu Val Ser Ser Thr Ser Ser  
20 25 30

Ile Tyr Ser Thr Pro Glu Glu Lys Cys Gln Ser Glu Ile Arg Lys Leu  
35 40 45

Arg Arg Glu Leu Asp Ala Ser Gln Glu Lys Val Ser Ala Leu Thr Thr  
50 55 60

Gln Leu Thr Ala Asn Ala His Leu Val Ala Ala Phe Glu Gln Ser Leu  
65 70 75 80

Gly Asn Met Thr Ile Arg Leu Gln Ser Leu Thr Met Thr Ala Glu Gln  
85 90 95

Lys	Asp	Ser	Glu	Leu	Asn	Glu	Leu	Arg	Lys	Thr	Ile	Glu	Leu	Leu	Lys
			100					105					110		
Lys	Gln	Asn	Ala	Ala	Ala	Gln	Ala	Ala	Ile	Asn	Gly	Val	Ile	Asn	Thr
			115				120					125			
Pro	Glu	Leu	Asn	Cys	Lys	Gly	Asn	Gly	Ser	Ala	Arg	Leu	Gln	Thr	Tyr
	130					135					140				
Ala	Ser	Ala	Ala	Thr	Leu	Leu	Arg	Gln	Cys	Leu	Gln	Tyr	Gln		
145					150					155					

```
<210> 37
<211> 7
<212> PRT
<213> mouse
```

<400> 37

Arg His Gln Pro Leu Lys Cys  
5

```
<210> 38
<211> 19
<212> PRT
<213> artificial Sequence
```

<220>  
<223> Description of the artificial Sequence: Peptide for production of AK

<400> 38

Glu Lys Gly Glu Asp Pro Glu Thr Arg Arg Met Arg Thr Val Lys Asn  
5 10 15

Ile Ala Asp

```
<210> 39
<211> 81369
<213> DNA
<214> Homo sapiens
```

<400> 39

gatcagactt tgaagagtgt ttgtaccatg ctaaagttta cagaatttat tcctgctctt 60  
tgaqqgtgca ttgcaaatcc aggctagagg gagagatacc agttaggana gtacagcaat 120

actctactgg gaaatggtga ggtgtttcgt gaagacaatg gcaacacaga tgaagacatg	180
cagatggagg aaataaagat ccagttgagc ttgttggcca gttggataga ggttgagggt	240
atgcatgatg gagcaatcta ggtttttgtc ttgggtaggt gtttccatga tagtactcag	300
aatgaatcat atagttgtac aggttgaatc ccacccatgt ttgcacaata gagtgactgt	360
ctagctgaaa tccagatgac actctgtatg ctaagctatg cttcatggaa ctgtataaag	420
gcacttgcta cataggctag tggcagatct ggaagtaacc tatatgggtat ataggaaatg	480
aggtggcttt tgtataaatc ctacagataa atttcatttc ctgatcctat tattttgact	540
catgttagcc caagaagagt attcagtact tcatatccct gaaggtaaga cagagtagta	600
ttagattcac tattttggcaa ataaaagggg tcaagtccta agatcaagct gatgaatcaa	660
cacctcatag gatattgtccc aaccaattat atggcttccc ctataaataa aatctagttc	720
tcttctctgg agaggaacag tgaagaatat cataacctat gctacaaact gcttgagtag	780
gagctacttc tctccaaggc tttatatcat tcattctggc aggccctct gtttgttctc	840
accagctcct gggaaaattta tttctcctct agtgatataa aagctctctg tttgagatga	900
agggctgccc agtttatcag atctgtatta gtctgttctc aggctgctaa taaagacata	960
cctgagactg agtaatttat gaaggaaaga ggtttaattg actcacagtt ccacatggct	1020
ggggaggcct cacaatcatg gcgaaagact aataaggagc aaagtcacat cttacatggc	1080
tgcagacaag agagcatgtg caggggaact gctctccata aaaccatcag atcttgtgag	1140
acttgttcac tattacaaga acaacagaca ggaaaaccgc cccctcaat tcaattacct	1200
gccactggga ccctcccaca acacatgggg attatgagag ctacaattca agatgagatt	1260
tgggtgggga taccgcaaaa ccatatgaag ttctttcttt gttactgggt accatatcca	1320
ttctgttgag gttctgagcc tttccagtta ctgtaactcc tctatctcct gtctgtgcta	1380
agactcagtg acctctctct gccttgcttc tgctttgtcc tgacctttc tgtgcatgca	1440
ctcactctag tttgcccacc tgaggtgaga gatgggtccag attagcaaca acaatctgtg	1500
gactaaaatc ctcttttaggg aggaagcaaaa attcagatgg atgttactaa acaaagctca	1560
gaaacagaga ccagggtgtg ggaagtaagg tagtagcctg agagcagctg gcagtgtttt	1620
agacctggag ggaggttagg tcatcagcaa tgaggagact gcctggaaaa tcctagaaaa	1680
ttaagacatc tggtcaggca aggtcatatc accagcacac ttcccttttc aagttgaatc	1740
cctttctctt gttaagagga ttcaagtgtc tttcttgcac tttgtcttct cttctatatc	1800

catgcttgca	atataaggag	acagcagttg	gctgtttgtg	ctagaaaaata	taaatggcca	1860
ttttgaaagc	atgccagaca	ggatctgcgg	caagttttca	atgttactgc	tgccatctgt	1920
tgttcttcag	tgctgggatg	tgaatctctt	ggcaaacatc	tctctaattc	tgaactatct	1980
ttcaccccca	tctagagata	ttcacttact	gaagtgcctt	tttaaagcaa	tgttcctcac	2040
caaggcgatg	ttctgaatgt	tttaaaatgg	aagaatctgg	aatgttttta	ttataataca	2100
ttttgtatat	cccaaagcaa	aaatcaattt	cttcatgggt	aatacttttg	taattttgtt	2160
tttaataata	ttttcctttt	aaatataaga	aatattttat	tgaattaata	ctttaatgta	2220
gctgtttcaa	gtaagataaa	acagaacaga	ttactgtttt	caaccttggt	cacagttagc	2280
tctgtaacta	agttgttgag	ctttatctaa	gcttttttat	ttttacataa	cgtttcctt	2340
ttcacttaac	cttgaaatta	tagtaatttg	ggaacttcta	ttcctctgaa	agagaaaagct	2400
aatgccaaag	atatttcaag	ggagaaagaa	ggttttttaa	aggagagaca	attcagctca	2460
gacttaatag	ctgtgattgc	tatttattaa	gcagaacgcc	tataactaaa	ttctcagata	2520
tccaaaaaac	agcctgtaca	ttctcaaaag	tgaagattac	acatttttcta	agttaaggta	2580
aaagttttgt	ctctgtagca	tcttactgat	ttctatcttc	tcattctgcc	ttaataatgt	2640
cactaaataa	atgtttgatg	cactaataca	tgaataaaaac	tattcatgggt	aatgattctt	2700
tagaaacaca	gctaagtttt	gtaattttgt	tttttaaaaa	ttaaaaattt	aaatataaaa	2760
atgtttttta	aaggcttgaa	tttcttgtta	aatgtacaca	ttttaagttg	taggctgtct	2820
ttaaaaataa	tctctccaca	cactgtagta	tttaaaacat	catgatatta	ctataaaaca	2880
tcaacaaata	gggcagtgga	aaacatggta	atcactaaaa	atgctcacat	gtcatatatt	2940
aagacttgat	aagtaaacca	caataataaa	tagaaaagaa	atagttgtct	aaaaagggat	3000
tctcaccttt	caaaccttac	cataaaaaatg	gaatataaaa	gaaggaagag	gaggagaaat	3060
caaattatat	cataaaattt	tctgggcaaa	aatattacag	aagaaaataa	gaaagattta	3120
tgagagttgac	tgaaacattt	ttgaatccta	tacataaaaa	tatcgttaat	taaaaggaaa	3180
aacaaagaaa	cagatttggg	aaatatttga	aactggtttt	tttttagcat	ttaaaaatgt	3240
aatacaaatg	gattatttaa	actccattgc	aaaaatacac	aaaggacatt	gacaatgtct	3300
ggaaataaaa	ttagctaagt	aagttataga	aaaactcagt	ctcacaattt	gacaaatgta	3360
actgaaaact	attaatataa	ttagtaacta	tttttacatg	tcaaaaattt	tgaattacta	3420
aaggaaacca	caatgcctga	aagtatccag	ggtttttttt	tttttttata	atattggcac	3480
tgtcatatgg	gtggcaggaa	ttgaagtgat	gttgtttctt	cagttattaa	gttgcatctg	3540

cagtgtttca aatgtccaaa acctgtgagt cagtaattct ctttttgtat atttatecta	3600
atacaataat tctaaacata atctcaatat atatgtacaa agttattcac tgcagtgtta	3660
cttacaatag ttagaaaatt gtaaaatgct ttatgcatct taaaatataa attgttgaat	3720
atataatagt ccatatgata taattatatac attattataa ataatgaatt agaaaataat	3780
ttaagagcat taaaataatt ataaggtaat atgaagtga tgaataatgt acagatacta	3840
taatcagcag agtgttaact aggtaaattt ttatgtgtgt atatactact tcctaaaaat	3900
gacttgacag aaatcatcaa aatgctaata gtggttactt ctgggtggga atacagatga	3960
tttactttgt tccttttatg tatttctgca ctgcccagtc ttccacagtg agcatatatt	4020
ggttttttaa tttatataag atggaaaaag ataccaaatg gtcttcaatg aatcctggag	4080
ttaactttca tgtgtgtcat atgttatatt cttaaacttat cacaaataga agactttaaa	4140
tcaacttgta cctatttcaa ctatataaca gcatctttaa aatgagcatt gaattaaact	4200
acccaaaacca accatcatga ggattattca agtaatgtgt ttaaacaaaa gaatttgtaa	4260
taaaattact ttatctcctt tgtgatttca gccattttaa aaaaaataga tgtttctact	4320
ctccttcaga tatcattaaa acataaactt gtgcctgact gcataaatcc cttttaaact	4380
aatatcactt attacgttta actaagtcta cctagggcct ccttggtataa agaacaagag	4440
ctttccattt tttgtttacc tagcccttct tgatgccacg acagaatagc tgtaaatctt	4500
cattatttat attctagaga aaataaaaagc aaataaaaag gtcagtgtat aaagtttatt	4560
ggttgttctc tttactcaaa acccacatgg tattaatggt agtctctatg aatatttcat	4620
ggataaaatc agagcattaa gtgcatacta aaaacaataa gaatggaaag actttaacct	4680
tatgtttata tgaattttcta ggttatcaag aagtttatag gctataggct ataaagtctt	4740
aggctatgat atagtaacct aatgtagact tccttgata catgaaaata atggtactaa	4800
gtacaaaacag aagatgagct taaaattatt ctttgagtcc tcttgatgga ttttttcccc	4860
cacactttcc ccaaaattgt tttatgccta tattgtagga gaccatgcaa gagacctaga	4920
gtctcttttt ctttcatcac tttccaatca acagcaaata ctatcatttt taccacaaaa	4980
tatatcttga aactcccttc ttttgattta cttgtaactc cccatcaaaa actgaagagt	5040
gtcacaatac ttcattaagt tccctacttg cactctacct ttaatataat tgtagcacta	5100
aaatgttttt aaaacatata tctgcttatg tcattttact gctcaatact atctgatttt	5160
ctattgcact tctaagatac tctaatttct tagcactcta tataaaatcc ttttaagggt	5220



tccctgctca ccttttcaga ctcagaacta tgtatttcct tttgcctgct gtacttgtac	5280
cactggattc ttgatttttg ttacttccag gtttttacac ttatttttac aataaatgtg	5340
aaataccctt tttgacaata tctacaaata tttcttattt gtctttattg ctctttcctg	5400
taatgtttag tcttcatttt cctgataatg gctatctaaa gttatctcct caaagaagca	5460
gttattttatt cacccaaattc ttctagtcct tctctggagt tttcttctca cttcattccc	5520
ttgggtttttg ccacaatttg taataatttg caatttggag tgttagaatg agggaataaa	5580
tcacaggtaa tgactatagt ttgtgactat gtaagattgg attcgttatt gatttattcc	5640
acaaacactg aggcaactgca tttagccaaa tgccaatctt gggcagtgag actctgaaag	5700
agaatctgct tccccacca taaactacaa agtgaaacaa ctcagaatgt acataaatta	5760
cagaatgaaa gcacactaga agtaaacaca gatgtggaag aggtaaagtg tccttgaaaa	5820
tcatggaaaag attcataaag ggaatgacat ttcaactgga ttctaaacca gttattcaag	5880
ctccacaagg ttgcacagta aatgagcagt ggcaggatga cataccttag aaagtaaaag	5940
gaatcttttt taaactgcta taaaaatcat tacatatata ttttgtaggt cgagagtaag	6000
gtatttaaca taaaaatcatt ttagtatatc agtgtttata tagacttagg tttttctcat	6060
ttaaaacctc ttttaatgac ttgtgctttt ctcatggta ataaaacatt ttcccaggaa	6120
gtgctgaata aatctttctt gaaatacggt ttattgcttt ctatcaatga ccctgaagta	6180
atacagaatt tacacttcag cggttgcaat gctcaaactt gacaggtaat gcactgtggt	6240
tgctgatata agaggatga tgtagggcta agtggttttg tgctcattta gctttcagga	6300
gaaaataatt gacttaacat tttgatacta aaacccaaag cctaacagtt aattcttggg	6360
attttaaatt attattgcaa agattattgt gccgaataat atgaaaatat tttatataat	6420
atttaaaaag tatatctctt tcttgggtatt atttaaatta ccataaaaaat gtgcgaaaaa	6480
gttatactga aatgtgatag gatcttttaa aagtggtgcc ttgattttgt taagtgttac	6540
ctagttttcc tctgaaaaca agaaacatac ccagaagttt tcacgaaatg gtctcatgaa	6600
tatctaagggt tagtccgtag tctcatctga gacaaggaaa gtcccttcca ctatgagcct	6660
gtaaaatcac aagcaagcta gttacttcct agatacaatg ggagtactgg tattgggtaa	6720
acacagctgt ttcaaattggg agaaattggc caaaattaat gggttacagg gcatgcaatt	6780
ccgaaatcca tctgggcagt caaattgtaa aactccaaaa tgatntcttt tgactccatg	6840
tntcacatcc aggacatgct gangcaagag atagggtccc ataatctttg gcagctctgc	6900
ccctgtggct ttgcagggta tatcaccctt cccagctgct ttcacaggct ggcattgagt	6960









gtaataaaaa acatctccca actcttttaa agctttggac caaatagcat cacagcctaa 13860  
 ttctaccaat catgcaaaga agaataccag tcttcttgat gctattacaa taaatcagag 13920  
 gaaggaattc tctctggctc attctacatg accagtgtca ccttgaaacc aaaacctgac 13980  
 aaggacacca caaaaagaaa actacaggcc aataaccatg atgaacacag atgcaaaaat 14040  
 cattaacaaa atactggcaa acggaatcca acagcacatc aaaaaataa tataccacaa 14100  
 tccagagggg ttgtatcaag gatacaagta tgactcaatg taaataaatc aataaacatg 14160  
 ataagcatct tcacagaata taagacaaat gaatatatga tcatctcaat agatgcagaa 14220  
 aaaaattttt gataaatttc aacatctctt catgaaaaaa atctctaaaa ctgagcatag 14280  
 aagaaacata cctcaatata ataaaggcca tatgtgacaa actcagagct aatatcatac 14340  
 agaatggggc aaagttaa aaactttcctc taagaactgg aacaagacaa ggatgcaaac 14400  
 tctcaccact cctatccaca tagtactaga agtcctagcc aaaacaatca gacaagcaaa 14460  
 agaaataaaa agtatctaaa ttgagaagag caagtaacat tgttcctctt tgctgatgat 14520  
 atgggttttgt atctggaaaa tactaaaaac tccagcaaaa acctcttaga tttgattaat 14580  
 taatttagta aagtttcagg atacaaaata aaaatacaaa agtcagtagc atttctatgc 14640  
 cccaataata aaatagctag gaaagaaatc aagaaagtga tcccatTTaa attagctaca 14700  
 aaaaattaaa atacctggga ataaatcaag gaagttaaag atctctgcac aaaactacaa 14760  
 aacactgatg aaagaaatta aggattaaac aaacaaattg agaaacatcc catgtttatg 14820  
 gatcaaaaaga attaatatca ttaaaatgac catacttccc aaagcaattt ccacattcaa 14880  
 tgcaatttct accaaattac caatgtcata tttcatagaa ttagaataat cctaaaatta 14940  
 gtatggaatg agaacagagc ccaaatagcc aaagcaattc tgaacataaa gaacaaatct 15000  
 ggtcctgact taatcactat gcaatctatg catgtaacaa aattgaacat ggattttatc 15060  
 aatttgtaca aataaaaaaa tgtaaaaaaa gaacaaagct ggaggctata gtagccaaaa 15120  
 cagcatggta tttttagaca aatggaatgg aatagaaagc tcagaaataa agccatatat 15180  
 atatattgtg tgtgtgtgtg tgtgtataca cacatacatg tatatataat gtgtacatat 15240  
 aatgttttct acatgttcta atatttatat tccattccat tatacatatt ccatttctgt 15300  
 atataggtta tatagaattg gaagactatc tgccattaaa aagaatgaaa tcctgtgatt 15360  
 tgcagcaaca tggttgaaac tggagtcat tatcttaagt gaaataatct aggcacaaaa 15420  
 agataaatat cacatgttct cacttatatg tgggagctaa taacttgatt acatgaaggt 15480



ccatttcatt	ccacctctga	ccctcccaa	atctcgtgtt	ctcacaattc	aaatacaatc	17280
atgcccttcc	aacagtcgcc	ccaaagtctt	aacacatttc	agtattaaca	caaaagtcca	17340
agtcctaaagt	ctaattctgag	acaaggcaag	tcccttctgc	ctatgagcct	gtaaattcga	17400
aagcaagtta	gctacttcct	agatacaata	gggtcacagt	cattgggtaa	atacacacat	17460
tccaaacggg	aggaattgac	caaaaccaag	gggtacagg	cctcatggag	gtccaaaatc	17520
caatagggcc	attgtttaa	cttaaagttt	caaaattatc	tcctttgact	tcatatctca	17580
cgtctaggtc	atgattatgc	aagaggtggg	ctcccacagc	tttgggcagc	tctgcctctg	17640
tggctttgca	gggtacagcc	ccactccagg	ctgcttttac	aagctagtgt	tgagtgcctg	17700
cagcttttcc	aggcacatgg	gtgcaagctg	taggtggatc	taccattctg	tggtctggag	17760
gatggtggcc	ttcatctcac	agatccacta	ggcagtaccc	cagtggggac	tctgtgtggg	17820
ggctctgac	ccacatttcc	cttccacact	gccctagcag	aggttcacca	tgagggtccc	17880
accctgcag	caaacttctg	cctgaacatc	caagcatttc	cttacatcct	ctggaatcta	17940
ggcggaggtt	tccagacctc	aattgttgac	ttctctgcaa	atgtaggctc	aacaccccat	18000
ggaagctggc	aaagcttggg	gctttcacct	tctgaagcca	tggccttagc	tgtaccttgg	18060
cccttattag	ttaaagctgg	agcagctggg	ttgcagggca	ccaagtccct	atggtgcata	18120
cagcaggggg	gccctggacc	cagcccacaa	aaccaatttt	ccctcctagg	cttctgggcc	18180
tgcgatgagt	agggttgcca	caaaactgtc	tgacatgcct	tggagacatt	ttccctattg	18240
tcttattaag	atttggtcca	tagttactta	tgcaaatttc	tgcagcaggc	ttgaatttct	18300
cctcagaaaa	tgagtttttc	ttttctatgg	catcatcagg	ttgcaaattt	ttaaaaacttt	18360
tatgctctgc	ttccctttta	caattaagtt	ccaattccaa	accatatctt	tctggataca	18420
taaaactgaa	tgtttataac	agcaccctaa	tcatatcctg	aacactttgc	ttctcagaaa	18480
tatcttctac	cagataccct	aaattatcgc	tctcaagttc	aaagtaccac	agatctctag	18540
ggcaggggca	aaatgccacc	agtctctttg	ctaaagcata	acaagagtca	cctttgctcc	18600
agttcccaac	aagtctctca	tctccatctg	agaccacctt	agcctggatt	tcattgtcca	18660
tatcattatc	agcatgttgg	tcaaagccat	tcaacaagtc	tctaggaagt	ttcaaacttt	18720
cccacatctt	cctatctttt	tctgaggcct	ccaaactgtt	ccaacttctg	cctgttacct	18780
agttgcaaag	ttactgccac	atttctgggt	atctttacag	cagtgcccc	ctcctggtac	18840
caatttacca	tatccattta	ttctcatgct	gataataaag	acatacccaa	ggctgggtag	18900



tttataaaga	aaaaagaggt	ttaattgact	cacagttcag	catggttggc	aaggcctcag	18960	
gaaacagaat	catggtggaa	gggaagcaaa	cacatcctcc	ttcacatggt	ggcagggaga	19020	
agaatgagca	aaacggggga	aaaaccctta	taaaatcatc	agatctcatg	agaactcact	19080	
ctcttgagaa	cagcatgagg	gtaacatgt	ccatgattcc	attacctccc	aacgggttcc	19140	
tcccatgaca	cgtgaagatt	atgggaacta	ctacaattca	agaggagatt	tgggtgggga	19200	
cacagccaaa	ccatgtcagt	catgatatga	gaaattatca	aattaagatg	tagggaaggt	19260	
ttttaaaaga	tttgagcaac	cacaaatgac	agatatgtgc	tatagtagtg	caaaatacca	19320	
ttttgctctt	attaaaaata	taattgttct	tgataatctg	aattataaat	gtcatggata	19380	
attatgatgc	attatgctct	cagcagctaa	aacttcaagc	aaaatacaca	cctagagagc	19440	
aatcagcctt	aacaataatt	ctataaattt	aattttcttt	atttctgata	attacatttt	19500	
agttgacttc	atatgtgatc	taaatacatt	accattattt	tggacttatg	atgtagctct	19560	
tgaagtacat	atatgatgta	gctcttaaag	tacatataga	agagcagata	aagtatcagt	19620	
tcaccatttc	tttgtagttt	gtgctttcat	gatgaatatt	ctcatcaatg	tacagattat	19680	
ttgcaggagc	cttttaaatc	catgtgtcca	ttttatgaga	cttagctttt	gtctgtatat	19740	
aatgtgttta	ttcagtgtgc	atggattaat	ttgagagagc	acaggtatgg	gtatctttac	19800	
agcagtgccc	cactcctggc	accaattttac	tgtattagtt	tattctcatg	ctactaataa	19860	
agactatata	tcacaataaa	ctgagaacca	gctggtaa	gagagaactg	tgggtccacct	19920	
tttcattgtg	gagttctcat	tttccttagc	ttatgctgct	tattcaacac	tatttctgca	19980	
taatctaata	cattcactaa	atgaagggtgc	tgtgttagcc	tccacatgat	attaatacag	20040	
cctatttaat	ttatccttct	ttagattaaa	aataaataag	tagtcatgtg	ccacagaatg	20100	
acacttcagt	catttgggtca	ttgaaggacc	acatctatta	ctgtgggtcca	ataagattat	20160	
aataacatat	ttttcctgta	catttttcatt	gttctgatat	gttttgatac	ataaatgctt	20220	
accatcgtgt	tagagttgcc	tgcagtattc	agtacagtaa	catgctgtac	acctaggagc	20280	
aacaggctat	accacatacc	ttaggtgtat	agttagggtta	taccatctag	gtttgtataa	20340	
gtacactcta	tgatgttctc	acaatgaaca	aaatcaccta	atgatgcatt	tctcaaaaca	20400	
tgtccctgtc	attaatacag	tatgtaacaa	tacagttagt	acaatatgta	atacatgact	20460	
atattcagaa	ttttagctat	ttctcttata	tttcaa	aatgg	attttcttat	gcactgtgtg	20520
gcacgggcat	ttcatttttag	taaccacagt	ctgggaaagg	agaagtcttt	gaaggatggt	20580	
gagcaagggt	atgacatggc	cagatgtgaa	tttttgatca	gtgactccat	gttagcagat	20640	





gaaaaaccag tttggacggt tcacagaatc aggatgatgt tgtgctgcat gttagctcaa	24120
agactaccct acaatatcgc agcttgcccc gcccttcaaa atccagcacc agtggcattc	24180
ctggccgagg aggccacaga tccagtacca gcagtattga ttccaacgtc agcagcaagt	24240
ctgctggggc caccacctcg aaactgagag aaccaactaa aattgggtca gggcgctcga	24300
gtcctgtcac cgtcaaccaa acagacaagg aaaaggaaaa agtagcagtc tcagattcag	24360
aaagtgtttc tttgtcaggt tcccccaaat ccagccccac ctctgccagc gcctgtggtg	24420
cacaaggtct caggcagcca ggatccaagt atccagatat tgcctcacc acatttcgaa	24480
ggtaaggatg tataaaatga tgctggaaaa atataaagga taaatatgtg ttagacacat	24540
acattacata taaatgtgtg tatatatata ttttaaatat gtataaggta tataatatat	24600
atatcttaga attcttttaa gtacacagtg agctctatga agcttatcat ataaacagct	24660
agcaaaaaaa atagttctca ttttgagaaa cagtcaaact tcaaagtttc actgtcattg	24720
tgatactagc aacacaaaca tctaagagac ttaaaagctg atggtaatac ctaagtgtag	24780
tgataaggca aagtaatagc ttgtaaaatt tctatagatt tccattcctc cttttcacat	24840
taaaaattaa aaccaaatag gttttcatga cttttggcat tcatttccag tgtcattttc	24900
ttgctggctc ttaatgagtt ggtgatcata aatgtagatg aagttgtttt ccttgtaaca	24960
gattccattg gacagattta tacagtgtca tatcttgaca cattaaagac aatcaagata	25020
tgacataatt tgaaactatt ccagtgtttg gtacagtatc acaactgaag agtgggctaa	25080
gctttctaac tcttcatctg ctttctttga catgactctg gtaaggatca tgacttggtt	25140
tctgttctcg gattgttttt ggtgttaa atgtgaagtt ctgctctaag atatcactgt	25200
ttttaaatac ccatgtgttt ttaagtggta ggaaaaataaa tgcagttaaa aattggggac	25260
aaatatctaa acctctctga gtctgttttc tcatctgcaa aatggtagag tgtgggttat	25320
agttcattat gggttcaata tttttaatgt ttgtttttat tctgttgact aaaccagaa	25380
ctttgatatc ttggaaagga aagattttga aacatttatt ttacaataaa gcaatttcag	25440
atacctgatt gtttgaaaaa cctaaaggct ttattcctcc gtagtaatat taatgctgca	25500
gaactgtctt tttaaaatac tgattctcat tgggaagaat gaattatggc gtatagggag	25560
agtaaatatt tctgtttctt aagtaaaagc caatagtgcc ctctgtggc ccattaccta	25620
tgaaacaatt tctcatattc gtcataaaat atttcactgt aggaaatatg gatttcattg	25680
caactcaatt agtaatcatt atgccattac ttcatatcat tgtatttcca tatttacata	25740







acactttgct tatccacatg ccagagatta taaaagaaat ccctaaactt gtccactta 30960  
 gttgttgata tcctcttcct gtatttttag agaggccatt tcttattttc tctagacata 31020  
 gcttttcatt ccttcttggt accaattgtg aattccttaa aatagagatg ataaaattta 31080  
 tagcctttta aataccta attatgatttc taaaagatgg tatagcttaa tttcattaaa 31140  
 atattcaa ataatgatact agaatacaatt aagttttaag caaacattca tatatctttc 31200  
 ttcacatgtg taaatgggaa ataaacatgc ctttttatta aaaataattt gaagacaaaa 31260  
 gataagtatt aaacaacggt ttataccatc tctgtcaatt ggaagttgtc actctaactt 31320  
 agccagagca gatctatctc attttgcagtg tgatatcata gcaaaagtct aatcagttgc 31380  
 ataggggaagg aaaaactaag atagtattta atcaatagga ttcagaggaa aattatgcta 31440  
 atgtgattta atctattttc tagtaatcct atcactaaac tgtcattgaa ttgtactgca 31500  
 ttagaaaagga actcaaata gtgtgacggc aatggacatc ttgtcacctt tagttggcct 31560  
 ttttcaatga gttaagcatt atatgtgtgt taccaaaaaa ttatttttta tagttcagag 31620  
 aaccattttt gttggatgtg taatttggaa gttttgttta cattatgtcc ttagggggtt 31680  
 tctttgtttt aacagcatgc agcttgacag aaatacacta cccaaaaagg gactaaggta 31740  
 tatattcctc tcagacaaat tgctacctct ctgttggtat gtaaactttg tgtgctgtct 31800  
 ctcttccttc tttgtttgtt tgcaatgtag cacatgacat tgaggacgaa atcactttta 31860  
 attttgatgg tttctctggc ccgaacagtt ggtgagatag ccccttaggt agagatacta 31920  
 gtagagattg aggctgtctc tcaaattaaa taaattccaa tgtgaatata actattttga 31980  
 agaaataata ctaaacaac aaacaacaa aacaaaaa aacaaacaaa aaacttgtcc 32040  
 caggcattac ttttttgggg gcagcaactt tggtagaatg cagaactcac ttcaacaaat 32100  
 taaaataaaa ttaactcttc taacttttgc ctattagagt catatgcatg caaatattca 32160  
 aaacccatgc agtctacaga tgtgggcagt taatgttgat aggttgaagg atgctacaat 32220  
 ctgaatcaaa gaaaacatat tttcatcatc acaggacaaa tgctgtaatt aaggtgtgat 32280  
 ttttatagaa tccttttgat aaaatctcaa aattgtttta atttctattt tgcaggggta 32340  
 ctgctatcag atcaatttaa atctgaatta atctaataat atttaataat ctcaaaataa 32400  
 ttattccatc cataataaaa aataaaataa aaatttaact tatggccatc ttttactgtg 32460  
 tacttttatc tgaggaagag atagaatgat ctactaatag aggtataaca ctgtatgtgt 32520  
 atgaaaagtt ggctaatttt ggtgctaaga atttacttac aaaaagaaaa agaataact 32580



tagtttggtg aaacactgaa taatggcgaa actaggtctt tctccattat tttttttctc	32640
tccaattttt cagcaatagc aaatagctgg caattattcc atgttaatat tttgatccag	32700
aaattttatgt tccagtaaag cgagcacatc tccctcctta tttttgtaat ctaggcatga	32760
tgtcaagtgg cagttaaaca aaagaactgt ttttccttta aaaaaaaaaa aaaaacaaaa	32820
gctgccaaata tgtattccat ttccctatgc cttctgtgac catccttcat ttcccttggc	32880
cctggcccac cactgtcctc catttgtagt ccatgttttc accctcttta catcctttct	32940
tgccctgtgc ttttgagttc tcaattaact tggctgtctg ctcattgctt atgatttcca	33000
actgcatatc tgatagaagc ataattttct cctcaaaacc ctttatctta ttttttttcc	33060
ctatgtgatt caaacagatg gcgtaagatc atctggaaga actgagcaat tataattaga	33120
ttcaatctgt ttgaaattgt tcattctgaa tagtaacctc ctctgaattg ttttcctgtc	33180
ctggcattgc cttgcccttg tagatgtgct taagtgtcat agctgtgctg ttttgcagat	33240
ataccccatc atctcggcag gccaaaccaag aagagggcaa agagtgggtg cgttctcatt	33300
ctactggagg gcttcaggac actggcaacc agtcacctct ggtttccctt tctgccatgt	33360
catcttctgc agctggaaaa taccactttt ctaacttggg taaaatattc taaaatattg	33420
attttgtttt gtttctttca ccaccactc tcacagaaac cctggaatct ctccataaca	33480
caacacgttt tcatttaaag ggagggataa aagcacttta acagtacctt tcatttgtgt	33540
cattgtttac tcttcacaga aaaatctcca aacattatgc tatttattgc tcatgacaaa	33600
tgcttaacat agattaatac tgtggttggt ttctagtcta ggctccagag gctcagaaag	33660
ttcacttgac ttgaaaaagt cttaccatta ctaagggttc aaggcagtaa ccagttcaga	33720
acatctgact ttaatcccag gggcctttcc attccattta agaatcctct taaaaaacag	33780
gaaggcatct ccttatttat ttgtctgaaa tattaataaca tccttaaaac aaaattagta	33840
atcttttgta gaaaatagaa acaattagga agaaaaaat atgtaattcc atgactcaaa	33900
gttaacttct tttaacactg ttaaagttaa aactccttaa aattcatata agaatttctg	33960
ttaagacaat actctgaaca ttttcaaata gatacaatga aaaataaatt accaacttag	34020
tcattggggt actttgtatt taacatcatt tgtatgaaat ataaaaatcat ttgcataaaa	34080
tttcattaaa agcactctga gtaacaaaat aattaaagaa aactaaacat gccagatacc	34140
atttaataga ttcaatgact ttaaaaaatat atttattttc tataaagtca catataaagt	34200
attttcatta tttttatggt aaatattttt attattagtt tatcagaaaa acttgtacat	34260
aaagatgagt attgatacat aatcttatta gagccagaga cgatcattcc ttctagaaaa	34320

acacatctct	gaatttagga	cggaggacaa	tgaacaaga	aatttcactt	tataatttac	34380
ctttgtcaaa	ctatcccaga	gcacatcaat	tccatcatga	aagtactctt	ttgacattat	34440
ataaaaaatt	agtaatagaa	aacacacaat	ccaaaacctt	atattttcta	aacttcaagt	34500
taatcatcaa	cttctcttag	atttttgaag	acctgaaaat	aaacataatt	tcaaataaca	34560
gaactcaaac	accatataca	tttgtaatga	ggcacaacag	tcaattttga	gccttgtatt	34620
ttccaggttt	tagctgaata	atcttcactg	ctttcttagc	tttttgccag	tctagtttgg	34680
ggactatttt	gccttactgg	gcctaaacag	agtgtaatat	taaaatatgt	taataagcca	34740
tactgagaat	aagataaatg	caggtttcta	actccttagg	gacacaagtg	gggacaacac	34800
attccatgaa	cacaggtgaa	tgaatgcccc	tagttttctt	gagttggaca	atttcatgcg	34860
atcatttttt	tctctgaggc	caaagtctct	ggtttgatct	tcctagcagc	ttccagaaca	34920
gaaagtgagt	ttactttgtc	tccatattct	ttttctccat	gctcggaat	ccctgcttt	34980
cctgatccca	ccacaaaaac	tcccctgagg	atgaagcctt	ggctttccag	gcttccaggg	35040
aagcctcgat	tcctggctgg	aggtagttgt	accacactcc	cagagggcta	aatcccataa	35100
acatcatctt	ctgtctttgt	agatcataga	actttttatt	atcatccagg	aagattttct	35160
ttttgaaaca	aggctggaaa	aactttatgt	cagtcctgac	ctgctcttta	atgactgcgt	35220
agagggagat	gccagctta	tccaacttgg	gttgacagaga	ggacagatct	gcagcccttc	35280
ttgccagaga	aaacatcctg	gcacagccac	aatcacaact	ccattcttct	ccgatagct	35340
cctttgcttt	gaaactcatt	ggttacttct	ccagtgtttt	caggtctata	ttctccaggt	35400
actccagcac	ctctttccag	ggcttggaca	aaaatacatc	tgtgttggcc	agcatcagtg	35460
ccaaggcagc	agcctccaag	ggctcctgca	cccatggacc	acatccacac	agagaagcac	35520
cttgggtcct	caagtgcctc	cctcttcttc	ccttctccca	aacctgaagc	ccagacacta	35580
aggggtcaaa	ccctcctggg	ccctgagggt	tccaagggcc	tcattacttt	ttcttttttt	35640
cactggaaaa	aaaattctaa	tcatgcacct	acagaagatt	gacatttttc	agtaagttgg	35700
actttccagc	tttcagccag	gacaagactc	aaggctatgt	cttttctatt	gcaacccttc	35760
ccactatatt	gagtagggct	tttagcaatt	gaaaacaatt	attttggcca	tggtttcata	35820
taagctaatt	atttcatatc	aaacaccaag	tttttgtttc	ctaacctata	tagtgataag	35880
agaatttacc	tataatgccca	aagaatgtat	agcttttatt	tgctttaaga	tgagttgat	35940
tttttaaaaa	agcgaaaagc	ctaacacttt	aacttcaaaa	aatgaattta	aatgtttgt	36000



ttcaaatcag gcatttggag atctttatgc tgtatggttt cagagtggaa aaaatgattg	37800
attcaaaaac ataataattta aagagttttt attgtattta cagttcacct gaacctctgt	37860
tcattgggca agaaaatgag tactcttaaa atgcaataat aaattaaagt tactttatta	37920
ttaaatttta aatatatata tatatactta ccttaaatat gtctcttgt tgtcttttag	37980
catcacccat ttttgatttg accattatct tttctgaata atcagtaaga tacaggatta	38040
ttattaatgt tcaaaagtgt cagtattcat gttttcttta ttctttctac caattaaaat	38100
gtgttaatat ataaaatttt tagaaatttt actataaaaa atcacaacat atattagaaa	38160
attaagatca ctacaatatg tcataattag tagactactg tgagctactg ccacagtaaa	38220
ctatggttcg tgtgtcgttc ccagcatgct agccctagta gaaaccattc ccattcaaga	38280
aagactaaca aagtatagct tacataaatc aaaaagtctt tggatgaaac ttcatattggg	38340
aaaataaccc aatcgctacc cttcaatttt ttatgaatga aaaaatggaa gaataaaggc	38400
ctctaagatc cattcaaagc caggagacac acaagaattt ctaaatagaa gagaaacaga	38460
agaggtcata gttcttgtga gccatctcat aacctgggtga gactcattgt catgcctcca	38520
tgcatgataa caatcgctca gattcatttt tcactctgcc acaagggtta catgcaggaa	38580
cattaatgtc aacctgtcac ttctaataat catctaatat tctctaaatt cgatggatcc	38640
ttttgcatat ggtgattggt aaacaccttt gcataggaac agtttctatg cttttgtact	38700
caaatcttcc tctaccttga atcctttccc atcttctgtg tcaaccttca atcttctcag	38760
aatgaactcc tgtcttctat tctttcggaa gcatagaatc tcacggtcag aagagaccac	38820
atctgggttca acccttcac tcttatgtaa aattttatga catctctagc ttcttcttta	38880
aaccaccaa tgacagaaac tactaaaatc tagaaataac acctttgaaa ttctttcttt	38940
aagagatcaa ataaaaatttt cctgaatctt cacctattgt tcctagttat atatatccag	39000
attctacaaa ataagtcaaa gttagattgc atatgacagc tcttcatatt taaaacaata	39060
taataaaactc actagttaat gtctagctgt agatgcaaaa gtagagagtg acttgggggtt	39120
atttaaaaac ccagtccagc cagacacatt ggatcatgcc tgtaatacca gcagcactca	39180
ggaggctggg gcaagaggat cccttgtcca ggagttacag gctacagtga gctatgatcg	39240
tggcactgca tactccagcc tggaagacag agtgagaccc tgtctcacia taatagtatt	39300
taataatatc ataaaaaccc agtccacatt tatataggat cctgttttcc tcaagttact	39360
acaaataaat atataatctt aataaaaagg tagtggtttt gccaaagatag tggcttggt	39420



aaatgtttat	tatacactat	ttccttataa	tctaactttg	ataataatga	tggtccta	aat	41220
catgaactta	catcaattaa	gagcttgaag	tgactgagag	tatttgctg	gaagcattta		41280
aagcccttct	tgggaaattt	agatgtttta	tattttactt	tctttttgat	tttgcttttt		41340
ccattaaagt	gattactatt	tttaaagaga	aaaccgaaaa	ctctagaaag	accatctttt		41400
cttcataaca	ggtagcagaa	aacaccatgt	tattacattt	ctagcaagag	cagtagaggt		41460
gacttgttgg	ttttgtgtac	tgttgcttta	gaaattgatg	taaggcttcc	cataaacgtg		41520
ccagaggaaa	agagggacgc	aatgggatct	gttattgaac	atttcagagg	cagactctta		41580
ccttaaatag	ggactcacta	tacattcatg	ttttcataag	tattgggatc	atgttcttac		41640
tttctatcaa	cctgctattt	tcatctttca	agcttaagag	taataggctc	tgtgtgtttt		41700
gtttttcagt	gagcccaaca	aatttgtctc	aatttaacct	tcccgggccc	agcatgatgc		41760
gctcaaacag	catcccagcc	caagactctt	ccttcgatct	ctatgatgac	tcccagcttt		41820
gtgggagtgc	cacttctctg	gaggaaagac	ctcgtgccat	cagtcattcg	ggctcattca		41880
gagacagcat	ggaagaaggt	aagcgttgag	ggggattaaa	gatgaagtca	ctttatttaa		41940
accctgagag	ggaaaccatc	gtgtcactca	catcaciaag	attcctgaag	aggaaaataa		42000
actagtgtaa	ttatcatttg	ggaaactaga	agcttgaaga	agttttattc	tgtattatct		42060
tctatttctt	tatgtatttg	gaaatatgcc	agaatttggt	tatattaata	cttggtgtga		42120
gaagagttta	gactaaatct	acttttccaa	tacagaaata	tacatataaa	ctattttccc		42180
aggtgcatca	aatatcagag	caaatgtttt	gtttgacatt	ttgggttaaag	agccataaag		42240
acacacaaaac	cagaaacatt	attttatgaa	aataccacat	gttgctgact	tttattccca		42300
ggaattccct	ctggtgctaa	ttttttatta	tatcatttta	gaattcatat	tgtacctact		42360
tttttgcttt	ataagtcact	atttcttcat	ccaatggcaa	taaaattgtc	acctaaccta		42420
ataaatatct	ttatagttat	atagttctat	gtaaatactc	caaataaatc	agcttgaaaa		42480
cctcaggaag	ctgagttgat	gctcaaatat	atatattttt	gtaaactgta	gaagctcaaa		42540
tgtcaaattt	aacaataatt	tgagagactt	ttctctttga	tttaatgaat	ttttttagta		42600
tccataaaga	aaacttacag	catacatatt	ataaagcatg	tcagctaagg	ataaaaataa		42660
actagacata	caaattcaaa	ctgattagaa	tgaaattatt	aaccctaata	attatgttta		42720
aaagaaaagt	ctccaaatct	tgagacatac	cagagtttaa	gtcttcagcc	atccatttac		42780
ttgtggtata	aacttaggca	agtttcttaa	ccttcttatac	cctaagttct	gcatctgtaa		42840



ggcttctgtc	taataacttg	gcaactgagg	tactttaggg	agtatggatg	gggctcttcc	44640
atgtctcaac	gtcctgactg	ccaaaaaatt	atagcaggct	ggttctcaga	atcttatagt	44700
tagttgttat	tacttaatth	ccctaaccac	ccgttcttta	ctttttctgt	aaaggctgga	44760
atthttgagt	agaccttatt	gttttaactc	tattgttctg	tttgttttct	ccagttcatg	44820
gctcttcatt	atcactgggt	tccagcactt	cttctcttta	ctctacagta	agtaatggct	44880
gttaagaaaa	agcttggtg	tttgccatgc	acacagatga	tgaaatagat	cattttactg	44940
tgaacagatc	acattcatct	atgacttgca	caggagtgtg	gtagcaaaat	aacggcatac	45000
tctaagctgc	ccaataccca	ataaagtgcc	aggtgctcca	cctgccattc	tttggtcact	45060
tacatgtgct	ttcacttggc	ttttgtgcac	tcatcataat	caatgagtgg	atgtagaatt	45120
cgatttcata	aaacctactg	aggtatgact	tggagtctct	gaaacctatg	atgtagtctg	45180
ctatactatc	atthttagta	tgacgagttg	tccatgtttt	gttctttgag	ccgtgactgt	45240
taattgttct	atagtattht	cttctcattt	tttatthttt	agthttattg	tgagaggatt	45300
atcgaagggt	aaaagcagta	agggtaaagg	gtaaaagcat	aaaagaacca	gagatgttht	45360
thtttaata	tacctthtga	aagagtgtga	thttthtaac	thttatthtt	atthttattht	45420
atthtttat	ttatthttt	ttgagtcgag	gtcttgctth	gtcaccag	ctggagtaca	45480
atgacacaa	catagctcac	tgcaaccttg	aactcctggg	ctcaagttat	ccttctgcct	45540
cagcctgtca	agcagctagg	actacaggca	cgcaccacca	tgcccagcta	atthtttaaat	45600
tgthtttag	acaaggctcat	tgctatatth	accagactga	tcaataccca	tggttcaag	45660
caattcctcc	tgctthtagc	tcccaagtg	ctgggattac	aggtgtaagc	cagcacactt	45720
agatagaaac	thttatthatt	aagagaaaaa	taccagtgtt	tcaagttctt	ttgcaaactg	45780
gtgacattat	aattcattth	tgacaaggag	agthtttctg	thtggtaaat	acaattctat	45840
ctthtttaaa	aaagtagcct	acaggaagtt	atathtttat	agtgagtctt	thtagagcta	45900
ggthtaacagt	gaggtatatt	taaaagcagc	ctactgaatc	tcaatgggac	ttgagtacta	45960
tgaataagcc	ttaatcctgt	actgtaaggt	tcatgaagag	thcatagcct	ctgctgtcac	46020
tgatcaactg	agcatcatgg	gcagtattht	thtctactcat	tatcattagg	thcaaagtth	46080
tgthttgaacc	thtctthtat	agattaatct	catatathtt	ctgccttaca	tagtcattca	46140
aaatctgact	gttatthggca	gaagtaatat	thttctaatc	tctcctthtca	atgattaata	46200
thacctatag	cttctagaaa	ttaagaaatc	acgattagtt	thtaggtaaa	tgtactthtt	46260



gtgcaaattg ataaagtgag gaatgtgtaa acacacatga aaaaaacaca taaaagaaat 46320  
 atattaagac ttagtggtcc tctgttgagg ccagcactgc catttggttg ggaattgtat 46380  
 tctgatttaa accattgccca ttacatcta tgtgtaacat caaaagatgt agcatcatta 46440  
 ttattctaaa tacatacaat aattaatatt tggataaagc taccttcatg aaacctaaga 46500  
 aaaactaaat taaaaagaaa gaaagaaaga aaaatacact tagatagaag aaataagggtc 46560  
 tagtgattgg tagcacaata gagtgactat agttaacaat aatttattgt acatttcaaa 46620  
 atagctagaa aagaagattt ggaatgttcc taacaggaag aaatgatatt ctctctaaat 46680  
 gaagaatggg atattccact tcccagatt tgatcggtac acagcatatg tttgtataat 46740  
 accacatgca ccccataaat acatacaact attgtgtatc ccaatattaa agattttttt 46800  
 gaaaaattta ttcctcaaga aaaggatcat gagtttaaga aaaaacagat tactagtcta 46860  
 ccagtgtcca gtagacctt ctgtgttaat aaaagtgttc tgtatctaca ctatctaata 46920  
 tagtaactat gaaccatatg ttgccattga ttatttgaag tatatctggc aaagagatga 46980  
 attgactttt ttattttaat taatttacat tgaaatagcc acatgtgctt agcagctact 47040  
 agattggata gtgcaagttt atagagaaca caaggggtac atttgtagat aggagtggga 47100  
 tgtcaaatg atgaggataa ttagaaagca tacatgagaa atattgtttt aagagtagaa 47160  
 tatgaaatgg gaacacagat taaaatagag tatgtatata tatacatata tatgtgtata 47220  
 tatatacata tgtatgtgta tatatatata tatatatgtg tgtgtgtata tatatatatt 47280  
 tataggccaa tatatggagg tagggatat cctagtgtta agtgagtaaa gaatggatta 47340  
 ggtgatcgag ccacatgaga aggtgatatt attagaaaat tgaaagttgt atttgagatg 47400  
 atgaaaatga tatatttgaa ttgaaaagta aactgtagta aaataattca aataaatgaa 47460  
 tatttgggga actacttaag agaaaaatca taaaacatga ggagtcattc tttccccagt 47520  
 ccgccatgat caggccttag gatttaattg gcaatgagaa aatacctatg aaaatgcttt 47580  
 ttaaactatc acatgaaaaa gcaatttatt attttcatg ccttcttaat aactctcaat 47640  
 agagatttag ttgatttgca ttttgcctg gttcaatcaa gaaattatcg cgtgacatca 47700  
 ggcaagttgc caaatttctt tggactatac ctataaaata aaatttgaaa atattagcta 47760  
 gatctaacc atttgtctcc ggatgtctgc aaagtgggtg gaaatcacia gcctaacctg 47820  
 atctgcagag gtgttacctt tggcaaaact atggtttttg tgtttgtttt gaaatctaag 47880  
 gccaaagcgc gtggctcatg ccggtaatct caacactttg ggaggctgag gcgggtggat 47940  
 cacttgaggt caggagtctg agaccagcct ggccaacatg gcaaaacccc gtctctacta 48000

aaaatacaga	aattagccccg	gtgtagtggc	atacgtctgt	aatcccagct	atttgggagg	48060
ctgaggcagg	agaatcgct	gaacctggga	ggctgaggct	gctgcagtga	gcgccactgc	48120
actccagcct	gggcgacaaa	gccaaacact	gtctcagaaa	aaaaaaaaaa	aaaaggaaaa	48180
gagggagagg	ggagggagag	ggagagggaa	tctaagccaa	cactgtgaaa	tattgtgaaa	48240
tatggagctt	ctacctaaaa	attcaaaatt	ttaaattcct	tttaaaaaata	attggaatat	48300
ctatggaata	tctagcaata	ctaagatgaa	attcctctgg	gttttcagtc	acctgtaatt	48360
gacaccttta	gatgttggca	tgggctctca	ggaagccaca	gcctccacca	atgcttttct	48420
tcctgacact	gaagctaaat	ttgggtggct	agttttcatt	gtgctgttgc	tttcctcatg	48480
ggaaagaaat	accctttgct	atztatattg	ctgtcaaattg	ggaaaatgaa	agacagccaa	48540
ggaagatcat	gtgactattt	aaatacttca	agtccattta	ttctttatta	gccttgtcct	48600
gttaggcatt	taaatTTTTg	atccctgcaa	tagatgtttt	ttgattaact	gtatattaaa	48660
aactatatTT	aacctgtttt	gaatttgaat	tctaaattgt	atTTTTtcat	gagagcaagt	48720
gtcattTTTTg	attcattgtg	gattgtttaa	catgttgctt	aacaaatagc	taatactaac	48780
gtcataactt	tttaattagt	aaatttgaat	ggataaatgg	ccacttattg	gcttatagaa	48840
taaataaaaa	cattttttatt	cagtcaagtg	tttcatattt	tttatcatct	ccaggacatt	48900
gggcttgctc	aaaaccattg	ttaaaaaaaa	aatggcaaatt	aatccagttc	catcatgata	48960
tcattaatcc	cacacctaag	ctactgaaaa	aaatatatta	atattctggc	tcattgcttt	49020
atTTTTtatgg	taacaccac	ctggtattaa	taaccacaga	gtacgaaaga	aggcaaaggt	49080
taaagcaaat	aatagTTTTg	aaaaattggg	agtgaaaaaa	gtcatgctat	acggtatgta	49140
tataatagat	atttaatgat	tatgcttgct	actagtatat	gtaacaggac	tattatagat	49200
taacaaaaat	gcggtgagta	tatttcttga	ttatttttta	aaagaataaa	ttattattta	49260
aaaatacatg	aattattttat	tgattcttga	atctttacca	gctttctata	attctaggaa	49320
gcctagaagc	agaattgggc	aggataaaact	ggcaaaaaat	gtaaaaagta	ggccgggcac	49380
ggtgggctac	agtgagtcgt	gaatgcgcag	tgcacctgag	tgatagatca	agatcctgtc	49440
tcaaaaaaaaa	aaaaaaaaaaa	aaaagaaaga	aagaaagaaa	aacaacaaca	aaaacaaaag	49500
caaagtacta	gggaaaacta	atagacatag	ttacatagtt	aattgtgcc	tatgttttaa	49560
ggcaatgaaa	cttttatctt	aatattcctt	gcttactttt	tattcaaaaa	ccaaactgtg	49620
tataaaacct	taaaattatt	aggatctaaa	aaataaaatc	tttccttaaa	aatctaaaat	49680

tgagatgtaa attattcaag agtgctttttt aaaacagttt tcttataaag gctattagga	49740
ttctaccact tagccacttt attatttagc cactatatta ctaagtttac atatttttaa	49800
aggtagtgaa aatatagggga agacaaagct cagggttaaaa gagtttctgg caaataaaat	49860
atatcctgat ggtagacta ctttgcttta tgttttctga aagaaaagca gtaaaaaaca	49920
gttcaggtag ttttgtgtca attaatctag aactatacca aaagtagaca tagaaaacga	49980
gagattgttt ttcagctttg gatctgctta tggcaataag cagacttgta ctattcaaca	50040
acattatgca ttcttcaact tttcccagaa taaggagct tcccaaatgc aatgggtgcac	50100
ataactcatt ttctggcatt ttgcagccca gcatgaagaa gaaaaacaga gctaggagtt	50160
ttctggaagt caagtcaaaa acaccctgca aattcctatg gcagtcctcc tttccataag	50220
ctgcatagcc aaaaatgttt gccagacact tttatcactg ggtgtttcag tgttttcatt	50280
gtttaagcgt tttgctgact tgtgataatt aaaattatta ataatcatta aagaaagaaa	50340
aagtagaagt aaataatggt aattatctgt gggtatcagt agaggctctgt atgttaccoc	50400
agctttattt gacattgttt gtgatcagta aatcacagaa taaaattctg acatctaaac	50460
cttggctaga ggtctctata attttatgga gtctgtttcc tacaatctgt atgaaagata	50520
cttcaatatt ttaagtttac atgcacccat cttttttaga gtataatttt ataactattt	50580
ggtttatgtt gcttatgatt tacatcttag agtcttttaa ttctgtcttt tgcttaaagg	50640
aatattatgg atcaaatgac ctatatttta agaatacctt atgggtttata tattaagaaa	50700
catttatata aaattctaaa gtaacttgct tgtactattt caattgaata acttaatgta	50760
tttcattcta ttcttctcat agtagataat aaaaagtaca tcatgattat tgtattcatt	50820
tatacttggtg gaattaattg aaaatagttt ttatagttaa agtctttctt tttattgttt	50880
tacaggctga agaaaaggct cattcagagg taaaaaaaaa tatgcaatat tttaatattt	50940
tctatttttag tttgcattca tgatgaaatt agtcttggtga ccactagagg gctctgtgat	51000
acaatagcag aactccacag gactgctgaa gtaaggcagc taattgataa atgggtctttg	51060
atattgcctc ttaaaaaataa aatgaaagga agtttgtata gcaagctgtc ctttcacatt	51120
ctagattgag tcttagctca acaccttaata agttttctat aatagtaagc actcattaag	51180
tcattgataa atgaaggctc atgggtcttcc tattttatta cagtcttttt cccactccct	51240
gtaagaccat ctacacagga taatgggttga aacttgggca ccaagcctcc acaacacagg	51300
atactagcat ctcagactat ctgttttgtg tcattatctt gttgcctcta actgccattt	51360
tatgtgtggt gtgtcaccta ttgttctaata cacatatctt acaaatacat atttgggtgc	51420

actcgtgagc	aaatcaaaact	gcatttcagga	aagaatacta	ttttaatttc	ccttggtaaa	51480
acattttgtcc	tgggtcaaaga	gagcaggagg	actttaatta	tgacttttatt	caagggtgagg	51540
taatggctgt	ttgattgggt	tacactgagg	caatcagaca	acagagaaaa	aaaatgcctt	51600
aacaacagct	tttgcaaaag	tatttccttc	ccttgaagtc	ttatttttatt	agccttttaaa	51660
aataaaaat	gtgctatggt	taaaaatatt	tgaaaattat	tgattaaacc	aattttgtctt	51720
tataatctct	gaaccaaaga	gtggatatga	tttttaaaaa	tcaaagtgggt	tttattttaca	51780
tcacatggac	atgacaaagc	ttctaacact	gatcatagta	tagctactga	agcatcgaaa	51840
tgctacatct	atttgcctta	gtagtagtta	ttcaactccc	ccttttatcat	tgatgctgta	51900
tcatgagtta	tgggtttaaaa	aaacaatttc	aatcacttta	cagtttcctg	gattatatatt	51960
taaagatact	ggaatcatgt	aatagagact	atttaatttg	agaaatgctc	tttgagtttg	52020
gattcattta	tgaataaaat	agacgctgta	ttttctgaaa	tcattcatag	tcattatctt	52080
ataaatgtaa	agcaaagtgt	atttttagact	ggggtgtatc	tgttccggaa	aaaaaaaaaa	52140
acaggaacga	agtagaatca	catttgggtga	aattatataa	gtgtctactg	tttccagctt	52200
agagttctct	actttgttag	agtgtttgag	ttgaccacca	tttatttttca	acaaaatcta	52260
atgcccgggg	caaaaactag	acagttaata	aactatgtca	agaattctct	ttcaaactga	52320
gacagcattc	caaaagttca	actacaacta	tagataagat	ttgtttttga	agaaatgaga	52380
agcatcaaaa	gtagaatgtt	taacatccaa	gtaactgaaa	tccttgaga	ctagatatat	52440
acttatagaa	cctagtgtca	gattgttata	aatgttctat	ccttattagt	cacaacatga	52500
gacttgcaga	acaaactgca	gaaagtgtct	gaattaaaac	tttaaactatg	atataatata	52560
tccttaccct	tttctgtttc	agttttattg	gagtgtgaac	ttaaactaaa	agaaagatac	52620
cttagaatat	acatttatatt	ggtttatcta	attagttgca	cctatcattg	gttttttccc	52680
ctgatttttta	agatgtggat	aagctataaa	gcctctctga	gctaataata	actcactaaa	52740
taaaggtctt	gataatacag	atttggaag	gcttctctgc	agtcattgaa	actccagcca	52800
ataacaat	aaatgtgaac	tgattaaatg	ttgaattaag	cccaagtttt	agtgattgca	52860
ggatattcca	tagcctttga	gaagttttca	aactatgaga	aattaaaatg	tacagaggaa	52920
aaaaaaacct	aagatttttct	gaaaaagaac	atggagtatc	ttttactaaa	aaagaacaag	52980
aaaaatatgt	gtgtatatatac	agttttttata	aagaaaatat	ttttctacag	ttttattacc	53040
acagttttttc	tagaaggaga	agaatcaata	cagagggtaa	actgctcttg	agtcatttgc	53100

catttgaggg atggcaaag gagcaagtga gcgactttg atttgtagat tagagtttga	53160
cacataacac tttgcttttg aatgacattt gcttgttact gtggagtcag tgttcatatc	53220
ctttattttc aggagttgct gctgatacaa tgggggttaga atgagctaaa tacagcattt	53280
gctttcttgg tttgaattct gggttttaag taaaaatcta cttgcctatt ccattgattt	53340
ttttaattgc attcagcaaa tccataaact gcggagagag ctggttgcat cacaagaaaa	53400
agttgctacc ctcacatctc agctttcagc aaatgtaagt cacttcattt ttaaaatata	53460
ttacaacaaa tttttataga ggaaaatgaa atcatttttag taacaaactt acaaattttc	53520
agtgcctgat acagacttag attaccaact agcaggactc ataaaaagtt aacatttttt	53580
gcctactcag taataaaatg taaatccaaa ctgatgagag gcagcaatat ggttaaaatg	53640
gcttggtgtt tctaataaga ttggaaacaa tagtaacagc catatgggtt acttcttttc	53700
ttgtttgcta tttttattac tctcttgc taagattccc tgacaatgta agaggggttg	53760
ttagtgtttg actttggaag ataaaatatt cctgtgccc gcctccttca tctcaatgta	53820
ttgaacaatt tgttaagcat ccagttaatt ctaaaatag aaattaggtc taaataggga	53880
tagcttagct gcactgtgga tgagatatgg tttgctcaaa aaaccttggc agccttctca	53940
tagcaattta aaaggttaca cttttactgg caccagagca gccaggatg gcagaaatga	54000
tgacaatgaa gaccgtcaat taaattaaca ttactgaat atcttccact gtgtcaggga	54060
gcactcagag tagatgcaga atgataaagg agaaatgtgg cactgttccc agtcctgagg	54120
agcaatggtg ttaagaacag cagtgagggg taaggaaatg cctgctattt tgccatatgt	54180
cttacctctc tcaactcaaca gtcctttgct cagttctgct gcatagcttt gggcctgctc	54240
tgtgcctccc caccctccc actgctcctc tactgagttt ttctatctcc tagacaaagc	54300
atgatatgtc aagagtgagc aggtgcagac ccacagtgt agacttgaat aagagccatt	54360
tttaaatttt ttttaagcta tcattgtgca atataaatc taagtatgtg tatcatttca	54420
ttcacaatgt attcatttta gcactgtatt tgaattgatt ttattttctg aaatttgga	54480
gaattaattt tggatttatt ctatttattt ttaatagatg gtgttaggag attcctgaaa	54540
ataatagcag tttttagata attgtttaag caatatgaga aaataagggt attatttaac	54600
cttggtgtgt ttttaagag atagtccaga ggcaaccgta aattttataa tataggctac	54660
atgtatagaa gtatgaaata ttgttgtcta ggttcctgaa tttgtacca gaggaagtag	54720
aataatgtaa atgtcagaac ctctgggtt gtgtttatct gcaataagaa aggtcfaatg	54780
gcaaacctta tttattagat tgtcaggata cttgcagatg tcttgaatga ttactcaggg	54840

tttcatttta	tttttaaatgt	cccttggttg	agctcatcat	ataattcaga	tattggaata	54900
ataaatggct	gctagacata	gtggaagatg	ggctgatact	ttccatttga	aatgtaatga	54960
tgcttattgt	cttcaaaaaga	aaaaactaaa	atggtatttc	acattttttt	gtttttgttt	55020
ttgttttttt	ttctctgaga	atctcattct	tactcatgat	tattggtttc	ttgtgtacca	55080
tttcaacatt	tttctattat	atgctaattg	gtatatatac	ttaatacaca	cgtgcaaaaag	55140
cttccacaca	cacacacaca	cacacacaca	cacacacaca	cacacataca	cacacatacg	55200
gaaccaaatt	ctaacatagg	ggaataatct	tcggagtgaa	ctctgtgctg	ctgtttgaaa	55260
atggagatat	aatttttagaa	aggttcctgc	agttggctac	ccacctcgtc	tgctctaatt	55320
atgcttgtca	cactattttc	actgatgtgt	tttcatgact	ttagggcatg	aattctcagc	55380
tgggtgttaa	tatgaccaac	aaagggtgaa	aacaggttct	tgcatttttt	taagtactct	55440
ttttatgtga	aaagcacaga	tatgcagata	atacataact	gaacatccag	catatctgtg	55500
gctttaaaat	atcacgaaga	agagcacaat	tagggaaaag	aaaacatcta	tagtgtttcc	55560
ctaggggaac	aatcatttaa	aaaaaaataa	aaataaggaa	cacagactag	aagcagcagt	55620
gccaaataga	taattcatgc	tagtctttgt	gttaatttaa	aaagtgctag	tcttggagac	55680
aaacgcccaa	attgctctag	gttccactca	gctgtatgtg	ttatcattag	tattaacttt	55740
tgcacgctga	tgggagactg	atatatatcc	tgttttatgt	tcctttaaac	aattttataat	55800
gtaattttaga	aaccttctca	aatcacatta	gatccacaca	aaaacctgta	catagcagct	55860
ttatttttta	atagccaaag	aaaggaaaca	accaaaaata	tcccttaata	ggccagttaa	55920
taaacaaatt	ctgatacatc	tatatcatgg	actactactc	agcaatataa	agaaatgact	55980
attgatacgt	gcatcaactt	gggtggatcc	caggggtatt	atgctgagtg	aaaaaagaca	56040
gttatagaag	gtcaaatttt	gtataattcc	atttatataa	cattccagaa	atggcaaaat	56100
taaagaaaca	gagaacagat	tagtgattgc	taagggctaa	ggatgaagga	gagagagagg	56160
tagtgtgact	ataggaagag	ggagatcttt	agttttgtat	tttgaatgag	atggccatca	56220
catgaatcca	catatgtcaa	tctattaatg	taaatcaata	ttgtattcct	ggctttgata	56280
tataatataa	ttttataaga	tatataatca	ttgggggaaa	ctggatgaag	gatacaaggg	56340
acctccctgt	actatctttg	caacttcttg	tgtatataat	tataaaatat	ataatgtatt	56400
aaaatgtata	aaataatatt	ttaagtatca	gatactgatc	tttactcagt	atatgaagtg	56460
ttctatcata	acgtaacatg	cttttctttt	atttgtggta	ttttagtttc	aaactaaaat	56520

ataaatcacc	taaagatcta	cgacagttct	tttgaaaaaa	aatcttgctt	ttaatttccc	56580
aggagtttca	accttaatcc	tctcttttagt	gtttctttat	ttggtagtga	tagggactat	56640
caaagcttct	taccatcaaa	tacatttact	gactaaaaat	agaaaaataa	tttacattgt	56700
aaaaatgtac	aaattgaatg	acagtcaaaa	ggtacaggta	atgaagatat	gcattaacat	56760
ctacttttaa	aaaaaagttt	attaaaaattc	tcttttagac	taatgcagta	tctgggaatt	56820
tatataaata	gatatgtata	taaatgacta	ttaaacaatt	ttaatgtcag	ttatatTTTA	56880
aacattttta	taatattggt	ataactatgg	gggtaaaaatt	ttgtatatat	ctgaacattt	56940
ttgttcttaa	ggaaataatc	atTTTTacat	atccaggaat	ttgaattact	ctcaagtcac	57000
ctattaatta	caagtcattt	tgaactcatt	cattttcttt	gtgtttgctt	tataatgtca	57060
ttttagattt	catgcatcat	aatcagccat	caaataattt	agttaatact	tgatttttcc	57120
tcagttgtaa	gaagtgcgtg	gtttaaattt	cattcagaat	gtttcatttc	atctgaatta	57180
atatctgtta	atgtatgtaa	tatacacata	tttttaacat	gcattgtactt	aaattgatta	57240
tagggacttg	gtaaaattac	ttatttatag	gatattttta	atataatcaa	ggatttttta	57300
aatctacagt	tcccatTTga	aagtaaaagt	aagtctttgt	ttactagttt	gttcacagta	57360
caagtaaact	ttctaccttt	tggttaaagt	tgagtgcagc	ccccacagtg	agaaattggt	57420
atattagaac	tctaatagct	ataatttata	gggatgaatt	tcaatgagtt	tggttctaag	57480
aaataatctg	ttggttttta	caacattttt	aagtatcaga	tattcatctt	tactcagtat	57540
gtgacatgta	ctctcatagc	ttacgtgctt	ttcctttatt	tggggtggtt	tttatatatt	57600
aattggtata	tgcataattt	aaacttgcca	taattacatt	tatatggact	ctaaacaata	57660
acttgtattt	taatttttta	atttgaaatg	catctatgtc	tctgttaaaa	tgcatTTctt	57720
tccttttgcc	caaatggggt	atggtaagtc	aagagagtct	ctagttagct	cacctctcat	57780
ttgactggca	gagtaaagcc	cttgttcagt	agaatgtgtg	ttaagccttc	cctccctttt	57840
gtaaagttgt	tctgaacaga	gctgcataaa	accacaggta	aagtgttaag	ctgattctac	57900
tagcatgtcc	ttagaaagga	gagcggttat	attggcaggt	cctattgcct	ggcgtttctg	57960
atcaataact	caccaacaaa	cagaaaacag	aagccgcaca	aggaaaggca	gaactaaata	58020
aatggtaata	gcaaacaata	agccagatag	cctctggcct	ctcgcccaca	ccttaaggca	58080
gctgggtcag	gtgggatgct	tttgtttgct	ttttaacgta	ttttctttac	aaatctcagc	58140
cattacataa	tttgaaatg	gacacaaggc	tagttattac	taacattttt	aaagacatta	58200
ctgaatgaat	gtgtaagaaa	acaaaaggct	ctttttgcct	ttcagcagat	aagtctttta	58260

acacaaaatc	tcttgggtat	tttgagattg	tgttctactt	ctttgcttat	ttaatatattt	58320
cataaaattt	gctagt tact	cttgcttttt	tgcattctctt	ctaagagaaa	acaattgggtg	58380
catattatta	atgagaaaaca	cttcagtggt	tggacaattt	tttgtagtgg	aaaagaaatg	58440
tgaaacttta	tgttgcagaa	tcattcttgg	ttcaactaac	tactaatttt	aaaacataaa	58500
gtcttaaata	tatataaagt	ttatatgggt	aaatatatat	tacatataat	atatgtttta	58560
tatttataca	taatatacta	tatat ttata	catgatatac	taaatat ttt	cccatataaa	58620
taataaaaatg	ctctaggcat	atatgtgtgt	gtgtgtatat	atgtatatat	atatatacct	58680
tcataacata	catatataaa	atactatatt	atatatactc	taggtataca	tatatgccta	58740
tatatgcacc	tatatatttt	tatattacta	tataatatat	agtatatatt	actatatata	58800
ctactatata	ttactatata	atatatagta	tatatatagt	atatattata	tagtaatata	58860
ttactatata	atatataaat	atatgtgtgt	atatatatat	atgcctagag	tgtttttaat	58920
ttgtcagtg	gctgtctctg	taatctatat	gaagaaataa	aatgtagacg	ttatgtataa	58980
tgatattttca	tcttgttgtg	tggcatcata	gtaattctct	ttacatatct	attcagatta	59040
cttttgcacc	agcctaatac	attgtatgat	tccaaaacca	aagagagtat	ggattgaaat	59100
gatattccct	ttactaatac	tcagtcctgt	ctattttatt	acctttatag	acttcaccta	59160
acacaagtca	ggggatattt	atcatcatat	taatacaatt	ttactctgac	cttaaaaatta	59220
tgcaactgct	aaaggaaaaa	tcagaaccaa	ataaactgtc	attaacaacc	cccctgaaaa	59280
tccatat ttt	ttaaaagtca	ttttatcaag	tctctcagac	aagatgtgat	accctataag	59340
tttaatcagt	tttactttcc	attttctctt	cattaagggtg	ataaagatta	tcattagtag	59400
aaaaattttc	ccttattttg	ctccttttcc	atttacccta	ttgagtgaga	aatttagcct	59460
ctcataaactt	ctaaaagtagc	aatgttaatc	tgataaaacta	aaccaagggtg	agataaaattt	59520
aagacaatat	tttttttctt	caacttttaa	gttctggcgt	acatgggcag	gatatgcagg	59580
tttgttacat	gggtcaacat	atgccatagt	gatttgctgc	acagatcaac	tcatcgcccta	59640
gatattaagc	ccaccatcca	ttagctattc	ttcctgattc	tctccctccc	ctaactccca	59700
ctgacaggcc	ctagtgtgtg	ttgttcccc	ccatgtgccc	acgtgttctc	atcgtttctac	59760
tcccacttat	aagtgagaag	aagtgggtgt	tggttttctc	ttcctgtgtt	agtttgctga	59820
ggataatggc	ttccagctcc	atccatgtcc	ctccaaagga	catgacctca	ttccttttta	59880
tagctgcata	gtattccatg	gtgtatatgt	accacatttt	ctttatccag	tttatcattg	59940



gcatttgggt	tgatttcatg	tctttgctat	tgtgactagt	gctgcagtga	acataatgca	60000
tgcaggtatc	tttataatag	aattattttat	attccttttg	gtatataccc	agtaatggga	60060
ttactgggtc	aattttctgct	tccagatcct	tgaggaaatca	tcacactgtc	ttccacattg	60120
gttgaactaa	tttactctcc	caccaacagt	gtaaaagcat	tcctttttct	ctgaaacctc	60180
tgcagcacct	gttattttct	gactttaata	atcaccattc	tgactgctgt	gagatgggat	60240
ctcattgtgg	ttttgatggt	accctttttt	ttatatgttt	gttggctgca	tgactgtctt	60300
cttgtaagtg	tctattcata	tcctgtctat	tcatgtcttt	gcccactttt	taatggggaa	60360
gtttgttttt	tacttgcgca	tttggtgaag	ttccttgtag	actctagata	ttagaccttt	60420
gtcaaagtga	tagattccac	aaatgttctc	ccattctgca	gattgtctgt	tcactctgat	60480
gatagtttct	tttgctatgc	tgaaggctct	taattagatc	ctatttgtca	acttttgctt	60540
ttgttgcaat	tgcttttgga	gtttttgtca	taaaatcttt	gcccttacct	atgtcttgaa	60600
taatatggcc	cagattttgt	tctaggggtt	ttatagtttt	tggattttac	ttgtaagtct	60660
ttaatccatc	ttgggttaat	ttttgtataa	ggataaagga	agtgggccag	ttttaatttt	60720
ctgtatatgg	ctagtcagtt	ctaccagcac	catttattaa	ttgttttttc	agtttcccca	60780
ttgcttggtt	ttgtcagggt	tgtcgaagat	cagatgggtg	taggtgtttt	tcactaacat	60840
aatcataaca	tacatttcat	tgaaaacaac	acgactcaaa	atgttcttta	gtaaccagtt	60900
ataagttttt	ttgtgcataa	ttacaaactg	ccatttcta	cataaacatt	ttgtgggttac	60960
ttatagctag	aaaatgtgag	taatatagtt	tatacagcat	actctttaca	atccccgattt	61020
ctttgtcaaa	ctttaattca	tattaaattg	ataaagtata	cacaaagggg	aaaggagagt	61080
aattttcttc	aagtttcaca	tttaaggatt	catagtagaa	tgattaaacc	ttacatttct	61140
ccactataag	gagaattaaa	atggaaatat	tgagtaaaat	cttacatttc	atttagtaag	61200
tgctaataaa	gggtttctgc	cataattttc	cttattttta	aagaaaacac	acaatttttag	61260
ttttagggtt	tagtaaccaa	ttttatgggc	atagtgggaa	tatttctaac	aggttaaact	61320
gaagtgacca	tcatgggcat	atatatatat	tttaaattca	catatatgaa	tactatacag	61380
taaaaactaa	cttatgtctac	ataccacatg	gatgaatctc	aaaacccatg	taaagcaaaa	61440
gaaaaccaca	aaagaatcat	gccatttgat	tacacttggg	tgggtttttta	aacaggcata	61500
tctaaacata	gtgcttttaa	gtgtaagctt	gggtaggaaa	aactataaag	aaaagcaaga	61560
aaataattac	cacagaagtt	atgtagaggt	tatctttggg	gaaggaagag	ggaataataa	61620
gaaggggaca	aaqaaqagct	tcttqgttct	tqaaatgtcc	tatttcttga	cttggctggg	61680

gaatgcatga	atgttcacta	tgtgataagt	cagggggctg	ttttcatttt	gttcactttt	61740
atatatgtgt	ggattttttc	acagttgaaa	aggtaaagt	caggtgtggt	ggctcacacc	61800
tataatccca	gccaacactt	tgcggggcca	agggtgggaag	aattacttga	ggctaggagt	61860
tggagagtaa	cccaggcaac	aggggtgaggc	actgtctcta	cagaaaatga	aaaaaaaaaa	61920
aaaaaagtag	ctgggcatgt	tggtacatgc	ctatagtctt	tgctacttgg	gaggctgagg	61980
caagaggatc	acttttagccc	aggagtttta	gcctgcagt	aactaggggt	gtggcactgc	62040
actccagcct	gggtggcagc	aagacactga	gtaaaagaat	aaaataaata	attaaaagtt	62100
aaaatatagg	aaaaaatgag	catagcctta	tgctaatttt	tcagttacta	ggctctgat	62160
catcacattc	cttgcttgtc	attgaaaatt	ttttaaaacta	tgatactttt	ttttagtgg	62220
atztatccaa	ttaaatctgc	taacaaaatt	ggtgtataaa	tctcaagggt	aagggtatgt	62280
ggagagtggg	tgtgtttgtg	tgagagagag	agagagaaga	gggggaggag	aaaaagaagg	62340
aagagggaag	gaatggaaaa	agataataaa	gagttgttct	gatagattaa	tcttttagtag	62400
atgtattccc	tacaaattgt	ttttctccat	attgcagtgt	caggtaaaga	aaggcatccc	62460
aggatgaatt	cagagctagg	aacatgcacc	tttgtatcat	aatgctaata	gaagggaacat	62520
gtacattcta	actgttacca	ataatggaat	atatttccgt	tattaagtaa	taagctttta	62580
ttctttgtat	ttttgtgatc	catttgatag	taggtgcctc	agcatttcca	ctctgctata	62640
agtacatgga	gatataat	atttaagtca	tcttattcat	gtctttcaaa	aagaaattca	62700
tttttgggcca	aggatttcca	aattttgccc	catatatagg	tatagtttat	tatagacttc	62760
gtttgcaaaa	tattaaatcc	ttatatcctt	ttaggacac	aataaaattt	tataagtttg	62820
agataatgta	cttgcagttc	tacctcaggc	cgtggtgaga	gattgaagt	cctcttcatt	62880
ttaacatttt	gggttcaagt	tgttgcataa	gggcatgcaa	atggaaactg	gcctattttt	62940
gagctttaat	aaaatcgta	aatacttctt	aatcttaaga	gttatagtta	tgtactacaa	63000
tatgtataat	tctctaatat	ttaaaacaaa	acctgaaagc	cacaaaagct	tactgtgaaa	63060
taaaatgtga	tggaatatta	tttctaactg	gcttacctgt	atttcttca	ttgaagggaa	63120
tatgaagtag	aaaagccctt	ttattgaaaa	gagtttgga	agtaaagata	actcttttca	63180
attcaattct	ttgtaagtag	aaaaagagta	aagataatgt	ttagctgtca	gcagatgtct	63240
gacacttgat	ggagcgtatc	attacaatag	agcagctaac	aatatctgca	aaggtcacat	63300
tgaaagtata	aaaatgagga	atatttgtcc	attgaccatt	tcagtgcact	ctttttgggg	63360

tttaagtcta	aaaatcttgg	cagatcagaa	ctttatatc	ggcattttga	gtgtcaaate	63420
tctacatgat	gtgcaagtca	gaaggagtta	ttacttgcaa	aataccatct	tctttcagaa	63480
gttaaaactca	cattaaatgc	caggagactg	aaacactgat	tttaagaaga	caaagtttag	63540
aaaagatgaa	tgaaaatgtg	tgttaaagaa	gagtcaccag	tcagagctaa	ctatgatagt	63600
catagtattt	aaagagttgg	aacacatgaa	attaagcatt	ttgtaaaatg	aaggcttttc	63660
atccatccac	ataagattct	gacatttaaa	ctatgtttct	tccattctgt	tcacaggctc	63720
accttgtagc	agcttttgaa	aagagcttag	ggaatatgac	tggccgattg	caaagtctaa	63780
ctatgacagc	ggaacaaaag	gtatgttcag	aaattgccac	tggagactga	aagaagacag	63840
caaattgcat	aggattctta	aataatacct	gaagctcctt	aaaaataata	ttccaggctg	63900
agtgcagagg	ctcatgcctg	taatctcacc	actttgggag	accaaggtgg	gtggatcact	63960
taaggtcagg	agttcgagac	cagcctggcc	aacgtggtaa	aatcccatct	ctactaaaaa	64020
cacacacaaa	aaattagctg	ggcatggtgg	cgggtacctg	taatcccagc	tacgcaggag	64080
gctgaggcag	gagaatcact	tgaacccagg	aggcagagga	cgcagtgagc	caagatcaca	64140
ccactgcact	ccagcctggg	agacagaaca	aaaaaaaagag	taataataat	aaaataatat	64200
tcaattctat	actaaattaa	aacaatgata	atacctttct	tttcagattt	taatttaaag	64260
attttatcag	tttactccat	attggaacac	acaaaaggcaa	acaaaatcct	tgctgggcag	64320
tctattaatt	tacttctgga	tggaactagt	aaaagaatac	tgaatgttaa	gaaagagaaa	64380
cagtcacata	agagaatatt	ctggggggcaa	actgttatgc	agttgacaag	aatcacactt	64440
tgataagaac	tttcacaaat	acatggtcac	taaatccagc	tatagggcat	ggctgtaggc	64500
taagacacac	aggaaggatg	cctgggactc	tgccaagtaa	gggacttcag	gttacagcag	64560
ctatgaaaca	aaggccaatc	ctgtgtaatt	ttgaaataac	agaactagt	tgccatctag	64620
ggatatcacc	tttgaagaaa	agtcatttgt	tatatcaaaa	tacttaaaat	gaacctaaag	64680
gattttatgg	tatgaaagaa	ggtataccaa	aaagaaagga	acggagaatt	tagttcacga	64740
agacaaatgt	attaaaaagg	tccatactgc	atagaaagcc	tggtcacctt	tcctgtgatg	64800
accagttagc	ttacttctct	gctgttagtc	cagtggcctt	aacttccttg	gataggtatc	64860
agagataggt	gaaacctata	gaattctatg	gagtgtgtgt	gtgtgtgtgt	gtgcgtgcgt	64920
gtgtgtgtgt	gtgtgtgtat	gaaaactgta	aatgtgcata	aatgatcagg	tgtccagagc	64980
tttcatctaa	ttctcaaaga	gaccattat	atcagaagtt	ttgggtattt	tcaagaatgc	65040
gttcctctat	ctatccatag	gaatggcttc	agttttgtct	ttagattctg	taagttatgt	65100

gattagctttt	acaaaagtag	tatgtattac	caaattttgt	cactttacaa	aagtttat	65160
ttaaaacaga	atgaatagtt	caatgaaatc	aaaagagtaa	atcgaatat	cttataattg	65220
ccaagtatta	ttagcacatt	gtattctctc	tcatattctc	cgtataccct	gcccgtgaga	65280
gagaatatta	tccattcctg	gaaaatctgt	tctagcacag	ctaacaaaact	ccttttgaaa	65340
cataaaat	cctttctt	ctccctccct	cctccttcc	ctcccttcc	tccttttcc	65400
ttttcttcc	ttccttctg	cctcttttct	atccttcc	tctcctcc	tacacctt	65460
cttccctt	ttccccct	gtctccct	ctttctttt	tgctgcagct	tgtcacttca	65520
ctatgtaata	taagaaccca	gcaaatagaa	ttagaaggct	ttttagagca	gctgacggga	65580
aagaataaaa	acactggccc	ccagtattct	tgaatgagaa	ttctggctat	gtctgttaaa	65640
agctgggtaa	tcttgagcaa	gtttatctaa	cctttcttga	acctcaaatt	caccttctta	65700
aaagtgggga	tgataatgac	tacctttag	gatcaccatg	aggagtaa	cagatactgt	65760
tatcatgtca	catgctaggg	gctacaaaa	aatattacct	tcctttacat	ttctctttt	65820
cccttgaaaa	ttataagata	acaccaaatt	cctcactggg	catataccaa	gcatattgtt	65880
ggaaatgagt	gttagaattt	aagtctcaat	atctttaata	agtcaaaaatt	aatagaattt	65940
ttgtctcca	cccaatattt	tcttgaactc	tggtatatct	gtaagtgaat	tttctcatag	66000
aaacatacag	agaattttct	catatacata	tagaaaaaaa	tgtagaggta	tgttaatgta	66060
taatgcctat	gattaatgcc	tgaatattta	aaaataattt	ctataacata	agagatttta	66120
taatgtgtct	acataatcct	taaaataaca	ttgccaaaat	tataaaattt	tctcagaaga	66180
tatcagaatg	tctcatattg	tccttatcac	ttttttaact	gaaaataaaa	tcacttcttt	66240
ttgaattgca	aactgtatac	acacaacaat	catgggttaac	tagtttatta	atttgagatt	66300
ataacttgcc	tattctcaaa	gtgatattta	aaagcctata	aaattatttg	caatgtgaaa	66360
tggtataatt	caaagacaga	atctaattaa	aaccagtaga	ataatgtata	taacaatata	66420
cctcagccta	gataattact	actgcaaggc	actgaaatga	attgaatttc	aaggaagcta	66480
tggtacaaaag	ggagattgtt	aggtgtgttt	tattctcatt	ttctgaccag	gagagcataa	66540
tttagactga	ggagaaaact	ctttggcact	aaattcaagg	acgaatttat	tgccaagggt	66600
tttaaattgg	ggcatggaa	taacaaaaga	caaaatcact	gttcaaatag	acatttctct	66660
aaaagctaag	ggcataacat	ttaatcatat	ttcactaaaag	gcatttcttc	agggagctga	66720
gataaaaggg	tatattgctc	tctggtgatt	caacaatcct	gagaaaaggc	ttgtgaagta	66780

tagagcagag attctttaaac tcccttcccc aagttataag tttcatttgt ctatatagtc	66840
attcatcaag tttatatattga atttgtgctc ttctaattgac aaaacagtac agacaatata	66900
gatatagaat gatagatata ggtctatatc tatagacata cctatctact agaactctaa	66960
aagcatatta tacatgtatg taatatccct catggagttt atattttctca tatatatctc	67020
atatatatgt atctctttat catggagttt atatttttagg aggtcacaga tgataataaa	67080
aatataatta aaacaggcca ggtgtggtga ctcacacgtg taatccctagc actttgaaag	67140
gccaaggcag gtggactccc tgagatcagg agttcaagac cagcctggcc aacatagtga	67200
aaccccatct ctactagaaa caaaaattag ccaggcctgg tgggtgggcac ctgtagtccc	67260
agctattcag gaggttgagg caggagaatc acttgaacct gggaggtgga ggttgcagta	67320
agccgaggtc atgccactgc actccagcct gggcaacaga gcaagactct gtctcaaaaa	67380
aaaatatata tatataatat atataatata tatataaata tatatattat ataatatata	67440
tataaattac atattttataa atatgttaatt tatatatata atatataatt aaaacatata	67500
ggatttcagg tgatgataag cactactgaa aaaagtaaag ctgagaatga ggatactgag	67560
aagctggttt ggaagctaaa acacaaaagta acaaaggcca aggtgggttac atgttcttga	67620
ttacatactt taaaaatgga taaactaaat taagactcag attctagtct ttgggcttca	67680
cagtgtgatt ttcagcaatc acatggcatt aatagcctga aactacatca aaattgtcat	67740
ttgatttata gacaaaaata actcccttga atagagaggg attcactcct aacacttttc	67800
ctatttccag atgccaaaata acacggaatc tcttgccaaa tttgtgtggc agaacactgg	67860
ttttatatac ttatagcctg gtaagaaaga aaagacatgt atgaataact tagaaggcag	67920
aaaattatca tgctattaga ctcagtacaa tgtcatgtgc attctcaaag gaaacatctg	67980
cagaggcagg agaattgctt gaaccctgga ggtgaagggt gcactgagct gagatcatgc	68040
cactgcactc cagcctgggt gacagagaga gactgcatct caaaaaaata aaaattacaa	68100
aaataaaaaa taaaaaatag tgatcaatct ggcagcattt tctgaaagtt aagcagtatt	68160
cccaatagct gctaaaagaa gacatgttat ataatactaa gtctgtaagt aggtaaaaat	68220
taagagaatt gttaatgtgc ttgctgggga gtgaaattat ctctaggcat taccctatac	68280
ctaacctagg actcagtaga ctatgatatt ggcgtagttt gaccaagaat tttatcctga	68340
tttcagatcg ttttctcttc accagcactt cttcaccagg attatatgaa aaaaattaaa	68400
cctgatgccc tgaggcatcc attatatgtg ctgaaataac ttcttttctc accatctaga	68460
atggtactag ctatgtacca ctcttgtcag aatcaaggaa attgtactc aaatcattgt	68520

gcagcttaat	tttctcacag	aaggccagtt	gagaaaggct	caacttctag	gaatccagca	68580
aactatattt	tttataagta	acattttttac	agaactactt	ctaaatcctt	gtgttcaa	68640
ttactaaagc	tatattcaca	gctaaatatt	tcagaattta	aaatttaaaa	gactttcaaa	68700
ttagtttccct	gtagctgtca	tgccaaggca	attagaacat	atgttaagggt	atgagggggtt	68760
tttcttgtta	gaaggtcaga	gcagggcaga	gaagtagccc	cttgtatgag	tgatgaagct	68820
cagatattga	ctcctatgct	aaccataaag	cctagtagtt	tgctcatttg	ttacctctct	68880
gaaacatttt	tttgggtgac	tacaaaacag	gaattgaaac	cttcaaaaata	agggaaatttg	68940
aaaccaaate	tttgaaaata	gataatgctg	caactaaaaa	tttagttgaa	taagattttt	69000
acattaactc	tccctaattt	acgttatgat	atttgccatc	tagaagtgtt	tttaaaaaat	69060
atattgctgg	agtcagatga	tgcatccatt	aatctttggg	gcatagaata	atgtgaatct	69120
aaaattttca	aattattttac	actactggta	tttgggtcaat	gtaattttatt	tgaaactaga	69180
tgcaataggg	atggccaggt	tatttcagta	gaacaactag	caagacttca	gatgcatgggt	69240
ggagtgggga	aaggaggacc	tgtttaagga	aactagagct	gggaagtgtg	agattaactt	69300
agtgccaatg	tgaggaccta	aaaagcagat	gtggtggaaa	atttaaacag	gcttgcctag	69360
aaggtcaagt	tagttgatga	cacttgatga	gattgtccca	agctttggga	ttctcaacaa	69420
agtctttggt	agtgagaaat	ttggaaagag	atcaggtata	gttaagaaac	tgggttggaa	69480
aggccaccag	gaaaggcgaa	tattctgaca	caaaatttga	tcatttttatt	tggaagcatt	69540
tcaagcctga	cctgaacgaa	ttgttttagcc	tcagatacat	gcataaaaact	gtgaaaagag	69600
acattgactc	aatttagctt	ctttaacatg	agaaactttc	gtggaaaact	agaactttac	69660
aagctcagct	ggtgttgggg	gcatcattat	cttgaatagc	tactggagg	aaaatgaaat	69720
cttagtttgg	ttctcagggt	ttaaaatata	tatcattttt	gaaaagtgtg	aagtaacaaa	69780
atatgatctg	attatcttat	tcctaaaaatc	ctttgcagaa	ttatcccagc	ctcaatcttc	69840
tcttttagtat	ttaatgagaa	taagaaactg	gaaatgactg	aattggaaga	gtagacttta	69900
aatccatata	ttgatggcat	atacattttt	cagttttttt	tctaaatgat	taatgaggat	69960
tctcaaaaact	tgagtatctt	ctatgtttcc	cttcaacata	aagaaattgt	atgaaaatat	70020
tttaaaaaatt	tctaattgatt	ttatagttag	ctatcttggg	aattcatttc	taatcatgta	70080
cctcatccaa	actccccact	atggacaaaa	ataaaaataaa	aattattagt	tgcatctgaa	70140
ggccacatta	caattttctat	gcattataga	aacctgagaa	aatgtatctt	aaaaaataaa	70200

tgtgaacaac taaccataat tatgaagaag aaaaatgaaa actagaaata aactattgaa	70260
aaatgtctat gtatcagtta agttttttatt ttaaaattct ttatgtttat ctctataata	70320
ctattgggaa agagagaaaag gaaaacctga ctttgttctc atccaaagga ggtgattcca	70380
ctgatttagc caaaataaga cttcctgggtt ataataaata ataaagtttt tgatgttttt	70440
tatatggtac cccactcact aggtgatcag acaccctcct gcaaaaaaaaa aaaaaatagc	70500
tatgcaataa agttaagtt ttatgtttatt ctttcaaggg gagaaacatc tgtttaacac	70560
agaccagaat atttcaacaa agtcatccca atatttatgg agatcataaa tcaagcgaaa	70620
aaatatattc atcaacaact aaacaaacta cattaaatag tctcaaagca cattttcact	70680
ttttttctga caggaaaaca ggtttcacaa gtgtggagac attttaccat ggcttttaac	70740
agtgaggaag gatgttttaa taaagggaaa aatttatatgg aaagctcaga gaaaagagat	70800
gggtgtggct tgagtgacaa ggtgagagca gatctcatta actgaaatga gagagaagga	70860
aggaattttg caaatatgga aagataacta gtgcaagttt gaacagatta tgtcaatcaa	70920
tgtagaattt ggctatcttt ttaatcaaag aagactatgg aatattttat aggtgtttgc	70980
ttatactcaa agtttttaaag aaataacagt atgaatttgg ttgaactaat ttttttcata	71040
gataggattc tcccaagtta tatagcatat atatttctta actagttatt cttcctttta	71100
catatattgt gccacattga gtaacaacta acctgctaag agctattggg ttttaaaaga	71160
taattaatat tagaaagtga tcattttttct gtttcatatt aaacatgata ttctgaaaaa	71220
gcaacattgc ctgaatgttc tacattttat ctttttgaaa acaggtttta taagagattt	71280
cttgtgaaaa gctgaacggt ctgacactga aataagtcag ctaactcaaa gctaagctta	71340
attttttgac actgttggca tgaggctca ttcccaattt tttcatttaa agccacaggc	71400
aaatgtttta acagattttta atccgtagta caagcattat tgatcttaaa ttttaaggata	71460
aaaacctgat tttaattaga atttaatatg cattctagta tttacgttgt ataattaata	71520
tttacattcc atgattccac tatgtaccat ttatttcttt ttgaataaat ttccagtagg	71580
agcagaataa attttcagtg aatattttat ttcttggggg atatttttaa atggaaaata	71640
tattaagttt cggtaaaatc tgttgctaag ttggcagtg acagaatata aaaattggag	71700
agactgagtc attatgatga attgggtctg acttttgtca tgacactgga aatttccac	71760
aaatattata ttcttctttt ataataaata tagtcgaaat gaattgcagt caagtatttg	71820
aagacccatc tataaattta ggcggttact gttgattttt cattatgaga gattcttcca	71880
ctcataagct actaaaagta cataaagaag gtctggttgt ttgttttaaa tgtgactggt	71940

ctctatcagg	aaaatgtcag	gtatccgatg	aaaatagata	tatgagggtgc	caggatatcta	72000
ttccaaactt	ggatatcact	tcaattagca	tcatcttttt	tttttttttaa	agtgtcctaag	72060
gttagaataag	tcaccagata	ttcccatgta	tgaagcaatt	ttctgcaaag	gccgctgtgg	72120
atgatctttt	taaaatatat	attctggggag	acattgagta	aagagaaaatt	atttaccaga	72180
gaatgaagaa	ccgaggcccg	attctttggc	tttctgccaa	agatgctgaa	ggcagtgatg	72240
aatgacaaat	acattaccaa	ggaattctcc	ctctaagagg	ctgacaaaaga	tctgattttt	72300
aggattatat	taccaccaag	aagatacccc	ttgtcactga	gcttctaata	gaaatatggg	72360
ctatactgaa	acaattctca	gttctttttc	tttctatctt	tttttgagtt	attttatctt	72420
ccaaaaatga	gttatttctg	ataaaaataat	tcacttaaat	aattatgaaa	gttcaaattt	72480
gtgcaaatat	ttttattggg	acatcttaaa	attactctaa	attcaaaaag	aaaatatatg	72540
ctttattaaa	atttgatctg	taagctgctt	tgtttgtaat	ttaactatta	tataaaaaatt	72600
gtataataca	tatatTTTTat	ttactttatt	cctgtgttgc	tttggttgg	tgagactagg	72660
tctccacatt	aggagtttta	ctgaatgaaa	aagtatcaga	atgtaacatg	actttgatat	72720
ggcatcagaa	tttaataaga	tgacatttaa	taggaattag	gggtaagttc	cagggttttac	72780
acttaaatac	aaataatcaa	ttttgcaggc	acaaaatact	tcaaacaaaa	tctgaaatca	72840
ttcatttgac	aaaacttcag	gtttgcagtt	gacaataaat	acaatacaat	gcaacagtgc	72900
aatagtgata	tctaaatatc	taatgtaatc	ataggtaata	ttagtaagtg	tgttatctga	72960
aatgagtggg	gtgatatect	gctttacttt	gtactggtga	gttctgggtg	ccacctttga	73020
aaggaataaa	gactattcat	atctctttta	taagacaata	agaaaaacaa	acaaacaaac	73080
aaacaaaaaa	ccacctctct	tacttttagct	gagaaagaag	ttattaggta	cagcttgaca	73140
agttcagcta	agcatccaaa	tcttccagga	ggttgttact	acataaaaatc	aaaccttttt	73200
aattcaacta	tgagcaggga	gattttattt	ttctttcggg	tactaaagct	tccaaactct	73260
gtttattcca	caggaatctg	aacttataga	actaagagaa	accattgaaa	tgctgaaggc	73320
tcagaattct	gctgcccagg	cggctattca	gggagcactg	aatggtccag	accatctctc	73380
caaagggtata	tttagaaatc	atttcatttc	caccaatat	aataggcatc	tattttattt	73440
attaattaca	gtagaactgc	atttactcag	tgtcactgtg	cattattaat	acatactagt	73500
tgtattaata	gttgtattaa	tacatactag	tagtattaat	acatactacg	ttggatttaa	73560
tgtgatcaga	atcctagaat	tttagaacag	tgacttccat	tatcagataa	tttttaaact	73620





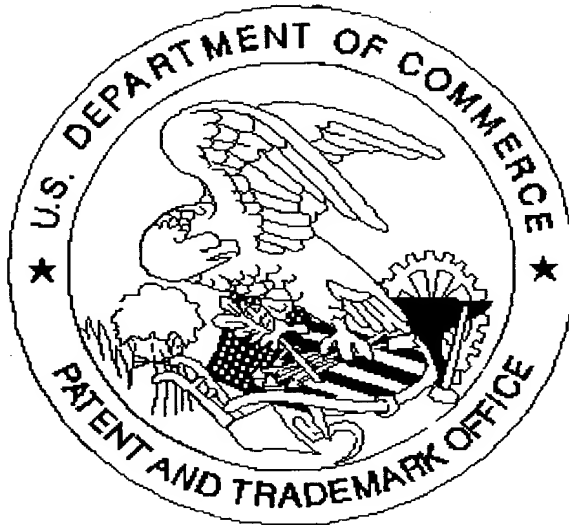
cctacattga atgacacccc ctgtgatatg tttctggaag cagcaatact tcccttctta	75420
aaattacatt atactttggg gcttttattt aaggatggtc tttcctgatt tacaatagta	75480
gagcttgttt tttcaccctt ttgaaagaca tcaagatgcc catgatgatg tcttgcatgt	75540
aacaggggtt tatttgaatt tttaaaagaa gaataaagta atttttaaat gaatttcaat	75600
ttaaatttta ggaaaacaat tatataaagt gagatatgct taaattgaag gacaaagtag	75660
ttctgtaggg gctacttctt tcaagacttt agcaactttc catgtggggg agtgatttat	75720
gtgatgcatg gaaaattact gcatatttaa agcttatctt agagctataa taaagcagct	75780
tatgttctaa atcttcatgt cgtaaatagg tccagaaggg atttaaaaag ccttaatcct	75840
tactttaaca cagcacaagt cactgaagtg aaacttgctg aaaggattcc ttttatgtta	75900
ggcaacaggt agctgaatat atctacagaa attgaaaaat tggaattctt ttgctcagaa	75960
atgtgggagg ggtggagctt aaggtaaaaa ataacagtta atatctaaat tgatcaagaa	76020
atatgaaaaa ataatttgct aggttttaaa actaacaaaa accatggtta taaaggtttg	76080
aatatatata ggatagttag attgtatttc tgtaatatta aaactcagca ttaaatttaa	76140
tgaacacaaa gtgattctta tcacattgac cattgacatt acatggaaaa aatagtcagt	76200
tggactaatt atgtgtcttt ccatgggtta ttaaggtaat tgtatggcat ataaatttat	76260
actggaaatc acattgaaat tcacttttag aggcccttaa aatatttctg taatatatat	76320
ttttaacata tgatcttaaa agatatattt ggaatgacac aacagtttta tagacaggcc	76380
tgactatcac acaaccacac accaatttgt gaatgtgttt ctatttcttc taaattaatg	76440
catcacattc attaacaaag tttgataaat gactatagtc tataataaaa tatttttgtt	76500
tacaaacata tttaaacacc tgctattaag tataggcatt atcagatctt aaaatacaaa	76560
gatttaaaaa attaccctgt ggtcatggag ctcaaatcc actgcaaaaa taatgtttgt	76620
gataagaaat ttgaaagttg aaggtaatag aaaattttac ctttattttt caaaatgtac	76680
cattgctttc taagtcacta cttctgtgta aatatggaat tgtttttctt taagatatac	76740
caaatatagt tggataacgc atgtattaaa attctgtcag cactaagttg ttttttagac	76800
atagtgatag gcaaacatag ttatattgaa tgaaaaatta gaatcaaatt tattaacac	76860
tgtgtactga ttgataccac atgccatatg cttgtatagc aatacaagggt ttggaattta	76920
taatggtaaa caaaatagat acggtctttg tctccataga acttttagtc tagtgggaga	76980
gcagaaggta aaggtaatgta tgtgatcatt ggtgaagctg aacatgtata cccaaacagt	77040

tataagttcc	aagatggaca	ataatgggtg	ccatagggaa	ggaggggtacc	aaggaaccta	77100
ctggaggtta	catagggaa	attattccaa	ggtagtaata	tttaagtga	tatccaagga	77160
ataattgtca	atcactttat	aagtactgag	ggaggagtat	ttcaaaagag	ctttgaggcg	77220
gaaaataaat	tagttccttt	atggaactaa	tgtaaggaaa	atactaagca	aacatgtaat	77280
aagaagaaca	cggttgatga	gttaagaact	gacaagatta	ctgaaggatt	gtaggccata	77340
tttagaagtt	ggatttttta	tctattctta	ttaaagtga	aagttattga	aagggtcttaa	77400
gtgggggagt	gatgatgaag	tttgcctttt	aaaaaagatt	tttctagcta	ttgtttatag	77460
aatggtttga	agatgaataa	gtccaatagc	tatacttgct	gtaaaggtta	tgttggtagc	77520
ttgaactggg	gcagtgggtga	cacagaggat	gggagatgga	aaatgacgag	tgaacaaaca	77580
catacctgaa	aatttaagtt	taaaaataga	cctctccatt	aattcagatt	gctgatattc	77640
attcggttag	ccattcttta	ctgaacttta	tgatgcccc	tatactgaat	taaatactta	77700
caagcactaa	aaaagaaatt	gttagggaac	agtaaaatgc	atttccttca	tttcacaata	77760
ttattaatat	tatggctttg	ctaactctta	ttggtgaatg	cagtcataat	tgaaggtaac	77820
tgatacttcc	aaggactact	tttgacctag	gattactatc	tttttaaaaa	tttagtatta	77880
aagaagtcaa	acacaattta	ttaattctgg	atataataaa	aattctgaaa	tactttaata	77940
ctttgtgctt	ttctatttgt	gaaagttaat	tattaggaac	gagctagcaa	atgctacttc	78000
tttttcaaaa	agctaatggc	caatcacagc	aaaaatttaa	agcactaaga	aatacctaca	78060
catattcttc	tattgcccac	ttatatgact	tccataatag	ttgattaaag	gataccggat	78120
tccttttattg	ttgaattaaa	acctcctaca	tgaaaacctt	gatttaggtt	tagaagttgg	78180
taatgttttg	gcatgcaaaa	ccagttaatg	ttctcatcat	tactttttta	aacaatgtta	78240
agagatgaat	tctagggatt	ataaaaaaaaa	aaaagctgta	tgtgttttctt	cctataaaaat	78300
ttttcagcat	gattgcctca	gtagaaaaat	taagggactt	attgatatat	atgtatatga	78360
aggtgaggat	acacatatat	acacacacat	atatatgtag	gtaaatacat	atattacatg	78420
tctatcaatc	catacatact	cattttattat	acgtttttgaa	agcaaccagt	tatagttttg	78480
ttgccatgga	tcattttttac	tattcagtaa	atcagtcaat	tgaagaggct	tgatttttatg	78540
gtattagttt	tttggaact	gtcagcttta	tagtaaattt	tgacatctta	caacttccac	78600
tgagattttt	ttgcttgact	aatctgcctt	gatgccata	agtatattaa	cggaaatgga	78660
ctaaaagcaa	atgtgacttg	aagcacaatt	ttgtaaattt	tcttagtgct	tcagtaatac	78720
ttaatactag	tgcattttag	gtaggaaaaat	tttcagtttg	ttttatttta	aataactata	78780

aatccttatag	ttgcttgtag	aaaagaaaaca	gataccttta	acatgattaa	atatcaaatg	78840
ctattctctt	caaaatatct	taactaaaga	agcactgcct	gctcttagaa	gttaagcaag	78900
gccataccat	atgctgcgta	catggctttt	aacacaatgg	atattagaaa	cagcctaagg	78960
ctgagcctgg	ctccactatt	tttcagctat	gtgaccatgt	gaaagttaca	tttagtaatt	79020
aaactcattt	cagtagtttg	ctttaagaat	aaaattaggt	actccggggg	catatcaagc	79080
atattgtaaa	acctagtttg	attattattt	gttattggta	ttactattac	tattctataa	79140
taagtcatgg	gcaggcagta	ggggtacatt	ggaagaattg	cactgtctta	aatatgtcct	79200
ctgtttaact	cacaaactca	gtctacctag	gctttctttg	gaggatctgc	ctttcattgg	79260
ctgtttgact	ttggccaagt	tacttaactt	cttttcactt	cagtttcctc	atctgtgaga	79320
ttatgtgctt	acatgacttc	aggttttgtt	ttggctctaa	tatggtatga	ttctatgaaa	79380
tggaaagtta	atacatttgg	ctctagtaac	tgtatttgaa	gcacaaatat	taaaaagcac	79440
aattaattct	cattctgagt	ttccatttac	tcttttaa	taatcattca	gaataaatca	79500
ttttggaaga	gctgcttgat	ccagggtattc	agtagaaatc	actagcatag	catttaattt	79560
tagacaaaac	tgagaactca	ttaaactgcc	agggtctatg	acttatatga	gattctcatt	79620
aaatcttaat	gtagataact	cagttaatta	aaacaaatat	ggttgactt	tattaaactt	79680
ctaaagtcaa	aactgcattg	aaattatctg	tacaaagcct	tgttgacctt	tattagagaa	79740
ctgcctctca	aaagacctaa	aagacttatt	tgttcagatc	gagactcttc	atgagccaat	79800
gtgatactct	ccctctattg	ctagatcttc	gcatcagaag	acagcattcc	tctgaaagtg	79860
tttctagtat	caacagtgcc	acaagccatt	ccagtattgg	cagtggtaat	gatgccgact	79920
ccaagaagaa	gaaaaagaaa	aactgggtaa	gttaccatcc	ttcatcta	tcagaagctt	79980
attaatgcat	aatgtgttag	gcctttttct	ttggggcttt	agtgatctgc	agtagtttac	80040
aaaggggtccc	attcaagcta	ctgagacctc	aaatgctgca	ctcatcacca	aaattggagt	80100
ggcatgtact	gaaaagcata	catttttaatg	ttgggactaa	acttgggttt	gaatcaccac	80160
tatatctaga	ccttttgagg	ggcctgaatt	ttctaacc	taaaaagaca	gttaatagca	80220
actatatttta	tttgtgaata	tcattttattc	acagatgtta	tctaattttt	ctatagtata	80280
actatacaaaa	ctatgtagta	taactataga	gttatactaa	agaaaaataa	gataacatct	80340
gtgaataaat	ggcttaaaat	aggggtttat	tgtgggcata	gagatgaagg	aaaagtga	80400
aaatgatgat	gatgggtgatg	atgatgggtga	tagtgggtctt	ggagggaag	gagaatggga	80460

gttaataaaag	ggaaagaata	aacaatgaaa	ctctcattcc	acctttggaa	tcgacagggc	80520
ttaccgtgtg	aatagtttca	ccctaaaaga	aatcaaccac	attagtgtct	gcttgatggt	80580
tttaaccaag	agaatatagc	agaaatatag	aatgcactt	taacagaact	gtaccttaag	80640
tttgctagt	atataattta	tgatattgat	caatagctaa	atagcccagg	ggaagatact	80700
gttactgcga	aaaattttaa	aacaatggag	tcaatgattt	cttttaatac	caaaaaaaaa	80760
atgtagattt	tgagtaaata	caactcttga	tgaaatccag	acataattat	cagaggattt	80820
tactggagtg	ctttctacaa	ataatgaaag	aaatatcttt	ttatcttaaa	aaatgtttat	80880
acaggtaata	ttttaaaata	ctgatcagcc	ttcattccct	tgatttgtaa	ttccacactc	80940
tttcatgttt	ctgcaagggt	aactctagag	gaagtgaggt	gaanataaac	cgtggacaat	81000
ttggcatgga	tntataaaaa	aaccctacct	tggcatgaat	gctatccatt	ttggcagtag	81060
gctttttatac	ctttttaaac	agattacctt	gtatgtcttt	tctttgtgtc	ttttcatttt	81120
aatctcaaat	tttaaagaga	tgtaaaacca	ctttctgaat	agagctgtag	gggataccaa	81180
ttctggtttt	gagtagtctg	gggttggaag	atttgaatag	aaaaatcaca	attaatgaag	81240
tgttaggtga	atttgatttc	attttgcttt	ttaagtttgt	actgtcagca	ggacatgact	81300
tgattgtagc	gctaaagtgg	ccatttaaaa	caaattgcct	tgaagagaga	agcattggga	81360
atggagatc						81369

United States Patent & Trademark Office  
Office of Initial Patent Examination -- Scanning Division



Application deficiencies found during scanning:

☐ Page(s) \_\_\_\_\_ of \_\_\_\_\_ were not present  
for scanning. (Document title)

☐ Page(s) \_\_\_\_\_ of \_\_\_\_\_ were not  
present  
for scanning. (Document title)

☐ *Scanned copy is best available. Some figures are very dark*